VOL. 15 No. 6, pp. 161-188 DECEMBER 1953

Philips Technical Review

DEALING WITH TECHNICAL PROBLEMS
RELATING TO THE PRODUCTS, PROCESSES AND INVESTIGATIONS OF
THE PHILIPS INDUSTRIES

EDITED BY THE RESEARCH LABORATORY OF N.V. PHILIPS' GLOEILAMPENFABRIEKEN, EINDHOVEN, NETHERLANDS

THE EQUIVALENT CIRCUIT OF A GAS DISCHARGE LAMP

by C. ZWIKKER *).

621.327.4.012

The gas discharge lamp, so far as its electrical behaviour is concerned, differs in many ways from other circuit elements, amongst other things in the non-linearity of its current-voltage characteristic. The behaviour of such a lamp in a circuit is, consequently, often difficult to predict in detail. It is sometimes considered that the gas discharge lamp can be replaced, schematically, by a simple resistance, but as discussed in this article, this may easily lead to false conclusions. Nevertheless for testing purposes, use may often be made of such a resistance, provided that an additional small self-inductance is connected in series. For the purposes of more exact calculations, however, it is necessary to employ a more comprehensive equivalent circuit which exactly reproduces the behaviour of the lamp.

The unambiguous representation of the electrical properties of a gas discharge lamp, for example, the behaviour of the lamp in an A.C. circuit, presents considerable difficulty. In the first place, the properties depend on the type of ballast resistor employed. It is thus incorrect to speak of the properties of the *lamp*. In the second place, all sorts of non-linear phenomena arise as a consequence of the fact that the relation between the current through the lamp and the voltage across it, i.e. the characteristic, is non-linear.

Expedients have, of course, been used to facilitate practical calculations for installations of gas discharge lamps. It is quite common, for example in the testing of ballasts, to consider a lamp as being replaceable by a simple resistance. It will be clear that this can give only a very rough approximation to the truth. Before going further into the considerations leading to a better equivalent circuit for a gas discharge lamp, we will first discuss briefly a few fundamental properties of a gas discharge [1]).

Fundamental properties of the gas discharge lamp

Under the influence of an electric discharge, a proportion of the gas atoms present in the lamp are split into positive ions and electrons.

Let us consider for a moment the behaviour of a single electron. The electron moves under the influence of the electrical field and collides with atoms and ions. During the time interval between two successive collisions it acquires a certain amount of energy. This energy increment is proportional to the field-strength F and also to the distance travelled, i.e. to the so-called mean free path, or more correctly, to the projection of the mean free path in the direction of the field. The mean free path is, however, inversely proportional to the number of particles per unit volume, and hence to the gas pressure p_0 at a given temperature, for example 0 °C. The increase in energy between two collisions is thus proportional to the ratio F/p_0.

The kinetic energy of the electron does not increase, however, without limit; the electron loses energy repeatedly in collisions with the atoms or with the wall. In the equilibrium state, the electron must lose on the average exactly as much energy per collision as it has acquired in its flight. The energy loss during the collision with an atom can take place in two ways. The collisions can be elastic; the atom then acquires part of the energy of the electron in the form of kinetic energy, i.e. the gas becomes warmer. The collision may, however, be inelastic; the atom then reaches an *excited state* or is perhaps even *ionized*. In the latter

*) Now Director of the National Aeronautical Research Institute, Amsterdam; previously with N.V. Philips, Eindhoven.

1) For a general survey of the fundamental processes in a gas discharge, see M. J. Druyvesteyn and F. M. Penning, Rev. Mod. Phys. 12, 87-174, 1940.

case — ionization — a new electron is released into the "plasma". The number of free electrons in the tube is kept up to a certain level in this way, since electrons are continually being lost by recombination. The energy of the colliding electron is, however, lost from the point of view of the purpose for which it is desired to use it, i.e. as radiation energy. This is because the fate of all ions is to be attracted to the negatively charged glass wall and to re-unite there with an electron. The energy thereby set free is lost to the glass wall as heat. The energy employed in the *excitation* of the atoms, however, does fulfil a useful purpose. The excited atoms always return, sooner or later, to the ground state, thereby emitting light or ultraviolet radiation.

The frequency of this process of excitation and emission is independent of the dimensions of the vessel. The rate of neutralization of the ions at the wall is, on the other hand, dependent on the tube diameter, and will occur more rapidly according as the tube is made narrower. It is then said that the wall losses are higher. Thus, to keep the number of electrons at a given level, relatively more ionizations must take place in a narrow tube than in a wide one. The average energy of the electrons in the first case must therefore be greater, i.e. the field-strength must be higher. The narrower the tube, the stronger the electric field and the higher the lamp voltage necessary to achieve an equilibrium condition, i.e. the arc voltage.

Whether a tube is to be considered as "narrow" or "wide" depends on whether the electrons and the ions undergo only a few or many collisions with neutral atoms before reaching the wall. It thus depends on the mean free path of these particles, or, more precisely, on the ratio of the tube radius R to the mean free path. In view of the fact that the mean free path, for ions as well as for electrons, is inversely proportional to the density, i.e. to p_0, the criterion of "narrow" or "wide" is the value of Rp_0, and this product is to be regarded as the decisive parameter in determining the magnitude of the wall losses.

The fact that in the equilibrium condition, the energy taken from the field must be equal to the energy lost by the electrons in collisions with the atoms and with the wall, may be expressed in the equation:

$$F/p_0 = \mathrm{f}(Rp_0) + S\,,$$

in which S represents the contribution made to this balance by the energy radiated in the form of light and heat. S may be considered to be a constant in so far as it is practically independent of the tube diameter and the pressure.

If, now, the lamp voltage is measured and by dividing this by the arc-length the field-strength F is determined, it is found that for a given gas or gas-mixture, at a given temperature, a functional relationship such as that indicated in the above equation, does in fact exist between the variables, p_0 and R. This relationship is depicted in fig. 1, as experimentally determined for the mixture of argon and mercury employed in low pressure fluorescent lamps ("TL" lamps) *). If the argon is omitted from the mixture, p_0 would be so small that the wall losses would become too large. The argon is therefore sometimes called the collision-gas, while the mercury vapour is termed the light-gas, owing to the radiation by the latter of ultraviolet of 2537 Å wavelength.

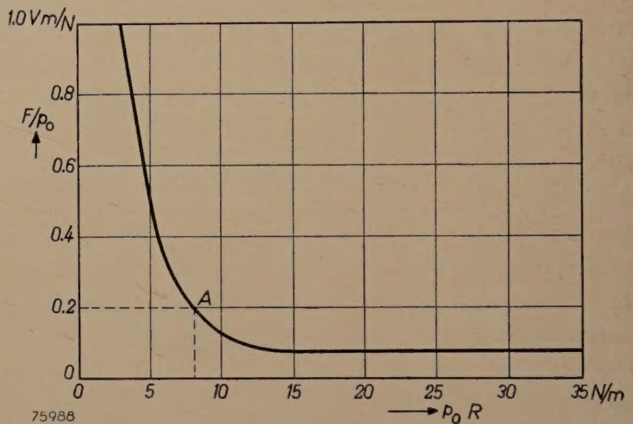


Fig. 1. Relation between the product of gas-pressure [2]) p_0 (reduced to 0 °C) and radius R of the lamp, and the ratio of field strength F to gas-pressure. This relationship is fixed for a given gas-mixture and temperature. The working point of the lamp is chosen preferably in the region of the point A.

The curve of *fig. 1* is useful in the calculation of lamp dimensions. The working-point of the lamp is chosen at the bend in the curve. There is no advantage in working much to the right of the of the bend, since the field-strength decreases only slightly with increasing R, i.e. with increasing weight of glass. A working-point much further to the left has the disadvantage that the small value of R gives rise to considerable wall losses. The working-point having been selected, any one of the three quantitites p, R or F may be varied according to choice, the remaining two then being fixed. The working-point [2]) can be for example (A in fig. 1): $p_0R = 8$ N/m, $F/p_0 = 0.2$ Vm/N; if we choose $R = 0.02$ m, then $p_0 = 400$ N/m^2, and $F = 80$ V/m.

*) In Gr. Britain "MCF/U"

[2]) The pressure is expressed in newton per m^2, the unit of pressure in the M.K.S. system: 1 N/m^2 (= 10 dynes/cm^2) is approximately 10^{-5} atmosphere; 133 N/m^2 is about 1 mm Hg.

Characteristic of a gas discharge lamp

The reader may have observed that in this treatment of the mechanism of the gas discharge the number of electrons has played no direct rôle. The number of moving electrons, i.e. the magnitude of the current passing through the lamp, is determined by the ballast connected in series with the lamp. This number has no influence on the field-strength, i.e. on the lamp voltage; a relationship has been derived above for F with which each electron must comply, without any influence being exerted by the other electrons. In this sense, therefore, the lamp voltage is independent of the current; the characteristic (current plotted as a function of voltage) of the gas discharge lamp is therefore a straight line parallel to the current axis.

Only on deeper consideration is it seen that the current exerts a small influence on the lamp voltage. At greater current strengths, i.e. at greater electron densities, the excited atoms have less time to return to the ground state before colliding again with an electron. There is thus a greater chance that they are brought into a higher excited level or even ionized by a second collision, even if the second colliding electron does not in itself possess the full ionization energy. The ionization is thus made easier, and more charged particles are produced. This shows itself in a lowering of the lamp voltage necessary to maintain the discharge. The characteristic thus exhibits a tendency to fall off slightly.

This shape of the characteristic has the important consequence that the lamp cannot burn in a stable manner without a ballast; this could only be possible with a sufficiently rising characteristic. It would be digressing too far to discuss the deeper origins of this phenomenon, and the reader is therefore referred in this connection to an article [3]) on the stability problem of gas discharges.

In the following we shall assume the necessity of a stabilizing element such as a resistor or choke, with or without a capacitor in combination. We shall, however, for convenience assume that the lamp voltage is independent of the current.

As already observed, the properties of a gas-discharge lamp depend on the ballast employed. This is also true of the characteristic. If, however, we wish to speak of *the* characteristic, we imply thereby the direct current characteristic; by the use of direct current, a resistor is the only form of ballast coming into consideration. The characteristic is, however, also dependent on the temperature of the surroundings, which determines to a considerable extent the lamp temperature, so that measurements may only be commenced after a period long enough to ensure that an equilibrium condition has been reached.

Operation on alternating-current; equivalent circuit

We will consider now the case where the lamp is supplied with alternating current at 50 cycles. The voltage across the lamp, v_{arc}, then varies to a first approximation, as shown in *fig. 2*; the voltage

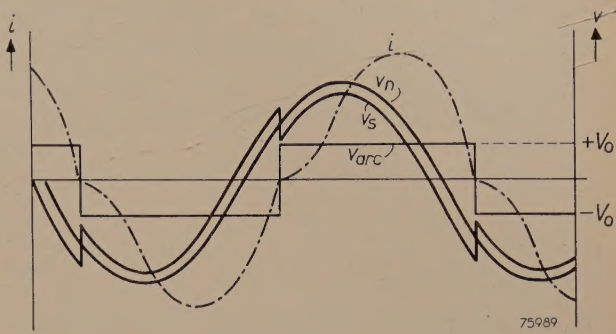


Fig. 2. Form taken by the various current and voltage curves in a gas discharge tube; v_n mains voltage, v_{arc} lamp voltage, which varies discontinuously between $-V_0$ and $+V_0$, v_s voltage across the choke, i current through the lamp.

changes sign each half-period, and is then constant during the half-period. We tacitly assume here that the time needed to reach the equilibrium condition is very small in comparison with a half-period. The fact that this is really not the case has certain consequences, for example in connection with the restriking of the arc at the change of polarity. This refinement is not important, however, for our further treatment. The lamp voltage changes discontinuously, then, during operation, from a positive to a negative value, and vice versa; let the amplitude of this rectangular waveform be V_0. The polarity reversals occur each time that the current i passes through zero (see fig. 2). We assume further, that the lamp is stabilized by a loss-free choke of constant inductance L. At the time t, then, if v_n represents the mains voltage:

$$v_n = v_{arc} + L\frac{di}{dt}. \qquad (1)$$

If v_n changes truly sinusoidally, the voltage across the choke, $v_s = L di/dt$, must display the same discontinuities, though in the opposite sense, as v_{arc}. This voltage is also shown in fig. 2. A discontinuity in di/dt implies, however, a kink in the curve of i plotted as a function of t. The i-curve thus contains kinks at the point where it cuts the horizontal axis. Such a current wave-form is in fact confirmed by measurements [4]). It deviates from a pure sinusoidal form, the current thus containing harmonics.

The behaviour of the lamp can be completely investigated by analysis of the solution of the

[3]) F. M. Penning, Phys. Z. 33, 816-822, 1932.

[4]) See, for example, the oscillograms shown in the article by E. D. Dorgelo, Philips tech. Rev. 2, 103-109, 1937.

differential equation (1). In this article, however, an alternative, more graphic method will be followed. The lamp in the circuit may be considered as the equivalent of an electromotive force, the value of which jumps discontinuously between $+V_0$ and $-V_0$. This waveform, sometimes termed a *signum* [5]) function (sgn ωt, in which ω represents the angular frequency of the mains voltage), can be analysed in a Fourier series, as follows:

$$v_{\mathrm{arc}} = V_0 \operatorname{sgn} \omega t = \frac{4}{\pi} V_0 (\sin \omega t + + \frac{1}{3} \sin 3\omega t + \frac{1}{5} \sin 5\omega t + \ldots). \quad (2)$$

The equivalent circuit of the lamp is thus constituted by a number of A.C. generators connected in series, having angular frequencies ω, 3ω, 5ω etc., and giving voltages of respectively: $V_1 = 4 V_0/\pi$, $V_3 = 4V_0/3\pi$, $V_5 = 4 V_0/5\pi$, etc. Since the voltage at the lamp terminals is considered to be independent of the current, these generators must be considered as being of zero internal resistance. Certain phase relationships also exist between the generators; each time that the total lamp voltage displays a discontinuity, all the partial voltages pass through zero.

In practice, the term containing sin $5\omega t$ and higher terms play no further rôle, partly because of the appearance of the coefficients $\frac{1}{5}$, $\frac{1}{7}$, ..., and partly because the signum function employed is only an idealised form of the voltage curve actually obtained. The latter has somewhat rounded corners, and therefore contains the higher harmonics to a lesser degree than expressed in formula (2). We will confine our attention, therefore, to the first two terms; the equivalent circuit of the lamp is thus reduced to two A.C. generators in series, with frequencies ω and 3ω.

From formula (2) it is seen that the second harmonic, and in general all even harmonics, do not appear in the lamp voltage. The fact that they do appear in practice, albeit to only a slight degree, is due to the asymmetry of the lamp (cathodes not exactly equal). The asymmetry becomes more pronounced as the lamp nears the end of its life.

Calculation of the current through the lamp

We shall now, on the basis of this equivalent circuit, make some simple calculations regarding the shape of the current curve in fig. 1, investigating particularly the fundamental and the third harmonic of this current. This is important in practice: it is desirable that the third harmonic is kept as small as possible in the lamp circuit, since such a distortion unfavourably affects the economy of the electricity supply system.

The fundamental

We can also make a Fourier analysis of the lamp current. The component i_1, with frequency ω, makes up by far the greatest part of the current through the lamp (as can be seen from fig. 2). In the first approximation we can neglect all the higher components, and thus omit the suffix 1. We assume further that the fundamental is in phase with the rectangular voltage v_{arc}. We shall discuss a better approximation at the end of this article.

The fundamental term i can be represented in the usual way, together with the mains voltage v_n and the component v_1 of the lamp voltage having frequency ω, in a vector diagram, since all these quantities vary with the same frequency ω. *Fig. 3* shows this vector diagram for the effective values of the various quantities, given in capital letters. The effective value V_1 of the voltage v_1, is put equal to aV_0; this follows from the peak value $4V_0/\pi$ (see formula (2)) by dividing by $\sqrt{2}$, so that obviously $a = 4/\pi\sqrt{2} = 0.90$. The voltage across the choke, $V_s = I\omega L$, is also shown.

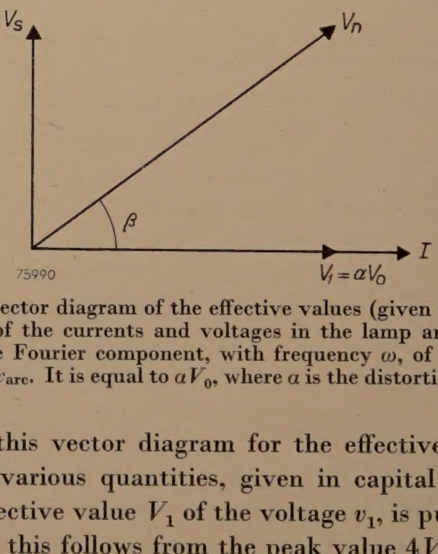


Fig. 3. Vector diagram of the effective values (given in capital letters) of the currents and voltages in the lamp and choke. V_1 is the Fourier component, with frequency ω, of the lamp voltage v_{arc}. It is equal to aV_0, where a is the distortion factor.

As assumed, the current I is, to a first approximation, in phase with V_1; the mains voltage is displaced in phase by an angle β in relation to the current. The voltage across the choke of course differs in phase by $\pi/2$ with reference to the current.

The power taken up by the lamp is (since I and V_1 are in phase), according to this approximation, equal to aV_0I. The quantity a is called the distortion factor of the lamp. It is less than unity because the current and voltage curves differ from each other in *shape* and not because of the existence of a phase difference (in this approximation zero phase difference is assumed).

[5]) The latinized form of *sign*. Strictly, the term applies only to a square waveform of amplitude *unity*. Thus sgn $\omega t = \pm 1$, alternating with frequency $\omega/2\pi$.

Although the only correct equivalent circuit for a gas discharge lamp is the above-mentioned series connection of voltage generators, it is very useful for practical purposes to have available a more convenient substitution element with which measurements can be carried out. This could be employed, for example, in the testing of ballasts.

On the basis of the above treatment, it is sometimes attempted to achieve this end by replacing the lamp by a simple resistance ϱ, of magnitude:

$$\varrho = aV_0/I. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

Such a resistance certainly dissipates the power aV_0I. It is clear, of course, that this treatment ignores the presence of harmonics in the lamp current. The value of the equivalent resistance depends on the current that is to be allowed to pass through the lamp: the term "equivalent resistance" must therefore be employed with some caution.

From fig. 3 we see that

$$I = \frac{aV_0}{\omega L} \tan \beta.$$

For various reasons, which will not be discussed further here, a value of 60° is chosen for β. This is obtained by a suitable choice of V_0, and thus of the lamp dimensions ($\cos \beta = aV_0/V_n$). Then:

$$I = \sqrt{3}\,\frac{aV_0}{\omega L}. \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

A phase difference of 60° also appears between the mains voltage and the mains current, which is here the same as the lamp current; this can be compensated by connecting a capacitor with capacity C across the mains (*fig. 4*). *Fig. 5* shows the vector diagram of the effective values of the currents and voltages so obtained. The capacitor must pass such a current that the combined current through capacitor and lamp is in phase with the mains. From fig. 5 it can be seen that, if $\beta = 60°$, the following equation must hold:

$$V_n\omega C = \frac{\sqrt{3}}{2} I, \quad . \quad . \quad . \quad . \quad . \quad (5)$$

or, according to (4):

$$\omega C = \frac{3}{4\omega L}. \quad . \quad . \quad . \quad . \quad . \quad . \quad (6)$$

The value of C can thus be calculated, either from the lamp current or from the inductance of the choke.

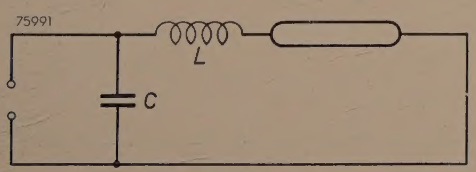


Fig. 4. Circuit for the compensation of phase displacement between mains voltage and lamp current. Compensation is achieved by suitable choice of the capacitance *C*.

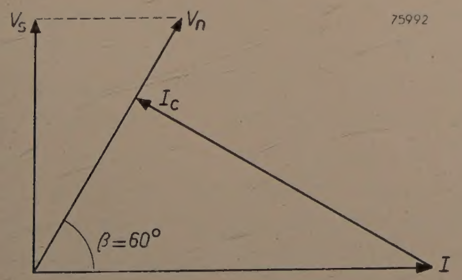


Fig. 5. Vector diagram corresponding to the circuit of fig. 4. I_c is the effective value of the current through the capacitor. I_c and I combined must be in phase with the mains voltage V_n.

The third harmonic

The equivalent circuit of the lamp also contains an e.m.f. of angular frequency 3ω, with a peak value of $4V_0/3\pi$, and this voltage generator also sends currents through the lamp and the circuit, but now with angular frequency 3ω. These currents are superimposed on the current due to the e.m.f. of angular frequency ω. There is a phase relationship which depends, however, on the ballast used; this phase relationship between the currents is determined by the already mentioned phase relationship between the voltages and the inductances and capacitances in the circuit.

Let us consider for a moment the simple case when no third harmonic is present in the mains voltage, an ideal self-inductance L again being taken as stabilizing element for the lamp. The effective value of the current component with angular frequency 3ω is then:

$$I_3 = {}^1/_2\sqrt{2}\cdot\frac{4}{3\pi}\,V_0/3\omega \mathrm{L} = \frac{a}{9}\,\frac{V_0}{\omega L}, \quad . \quad . \quad . \quad (7)$$

in which the distortion factor a is again introduced.

Comparison with (4) gives for the ratio of third harmonic to fundamental current, for this case:

$$I_3/I = \frac{1}{9\sqrt{3}} \approx 0.06\,. \quad . \quad . \quad . \quad . \quad (8)$$

Thus the lamp current contains about 6% third harmonic. The method employed, i.e., the treatment of the lamp current as being almost entirely due to the fundamental term, thus appears to be admissible.

Influence of the lamp circuit on the third harmonic

The extent to which a gas discharge lamp produces a third harmonic in the mains current, i.e. causes distortion, evidently depends considerably

on the circuit employed. A few examples of this dependence are given here.

A capacitive ballast is often employed in place of an inductive ballast. The former comprises a choke of inductance L, in series with a capacitor of such a capacity C that its reactance at the frequency ω is equal to $-2j\omega L$, thus $C = 1/2\,\omega^2 L$. The absolute value of the total impedance to the fundamental current is then again ωL, but the impedance is now capacitive. The advantage of such a circuit is quite clear. In a circuit containing a number of lamps, some of which are connected capacitively and others inductively, it is possible, without using correction capacitors, to reduce the phase difference between mains voltage and fundamental current to zero.

The numerical value of the impedance of the series-coupled C and L to the third harmonic is equal to:

$$|Z_3| = 3\omega L - \frac{1}{3\omega C} = 7\omega L/3$$

and has an inductive character. The third harmonic current in this case amounts to:

$$I_3 = {}^1/_2\,\sqrt{2}\cdot\frac{4}{3\pi}\cdot\frac{V_0}{7\omega L/3} = \frac{1}{7\sqrt{3}}\,I,$$

i.e. to about 8% of the fundamental current.

A special case of the above-mentioned combination of capacitive and inductive elements is the so-called "tulamp" circuit (*fig. 6*). Here also, due to the value of C chosen, the fundamental mains current is in phase with the mains voltage. Since the fundamental currents in the two circuit branches differ in phase by $2\pi/3$, the third harmonic in the two branches are in phase. No reduction in the third harmonic component of the mains current is therefore obtained by the use of this circuit.

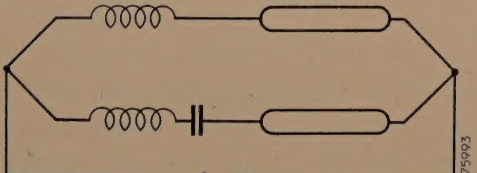


Fig. 6. "Tulamp" circuit. The value of the inductance is the same in each branch, and the capacity is so chosen that the impedance of the capacitive branch is equal and opposite to that of the inductive branch. The total current through the combined circuit is thus brought into phase with the mains voltage.

It is, however, possible, by use of a different type of circuit, to eliminate the third harmonic in the mains current, even when only one lamp is present. If, in the circuit of *fig. 7*, we so choose the capacitor and choke that $3\omega C = 1/3\omega L$, then no current component of frequency 3ω appears in the mains. The mains are thus, for this frequency, short-circuited. The third harmonic flows through the lamp, however, undiminished.

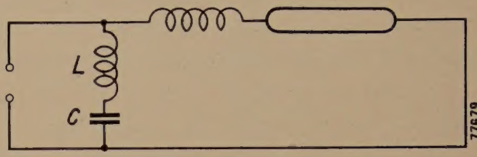


Fig. 7. Circuit with capacitor and choke chosen so that $3\omega C = {}^1/_3\omega L$. With this circuit the mains are short-circuited for currents of frequency 3ω, but the full third harmonic passes through the lamp.

A final example of the influence exerted by the circuit on the extent of the distortion in the current, is provided by the star-connected circuit of *fig. 8*. The three fundamental currents are separated $2\pi/3$ in phase, and the third harmonics will be in phase if a neutral conductor is present. In that case a three-fold third harmonic passes through the neutral. In the absence of a neutral, however, the third harmonics can have no finite values; no current of frequency 3ω can then pass through the lamps.

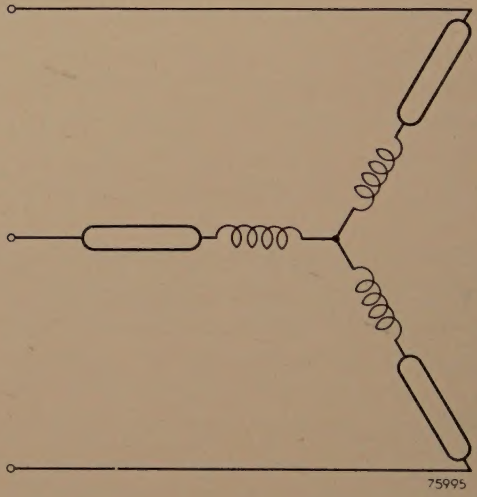


Fig. 8. Star-circuit for three "TL" lamps with their associated ballasts.

If the mains voltage itself contains a third harmonic — and that is usually the case on week-days — all sorts of conditions arise. It is therefore never certain that the third harmonic, such as is obtained under ideal conditions with a given circuit, will be produced in practice even with the same circuit.

Distortion produced outside the lamp

The important question in connection with distortion, i.e. the amount of third harmonic which a lamp produces in its circuits, is still only partly answered. In order to be able to estimate the increase in third harmonic in the mains caused by a given installation of gas discharge lamps and ballasts, we must take into consideration also the distortion caused by the ballasts — i.e. the chokes — themselves. This arises owing to the fact that the iron in chokes always exhibits a curved B-H characteristic — the magnetic induction B saturates when the field-strength H, and hence the current through the choke, reaches a certain value. The deviation from the ideal linear characteristic is greatest when the current through lamp and choke is at peak. The magnetic induction, therefore, does not vary in a pure sinusoidal manner with time, even if the current itself were purely sinusoidal. The third harmonic i'_3 produced in the choke, is proportional to the derivative of the third harmonic in the Fourier expansion of the variation of the magnetic induction with time. Since this derivative is of course zero when B itself is at maximum, i'_3 will be exactly zero when the current i through lamp and choke is at maximum. As we have seen, the third harmonic i_3 produced by the lamp is then just at its maximum value. The total third harmonic current is then, to a close approximation, equal to $\sqrt{i_3^2 + i'^2_3}$.

We have already seen that i_3 normally amounts to about 6-8% of the fundamental current i. There is no advantage in so constructing the choke (by choosing $B_{\max}$ so low, which means making the choke very large), that i'_3 is much less than 7% of i. If i'_3 is, for example, made *equal* to i_3, the total third harmonic is still only 10% of the fundamental current. This percentage is, in fact, found in the usual installations.

Improved equivalent circuit

We will consider somewhat more closely, in conclusion, the case of an ideal choke employed as ballast. *Fig. 9* shows the current component i_1 of frequency ω, and the third harmonic i_3 passing through the lamp, together with their phase relationships. If we sum i_1 and i_3 to give i, then a current waveform is obtained similar to that shown in fig. 2, particularly in that the maximum lies to the right of the centre.

If we consider the diagram more closely, we see that the current $i = i_1 + i_3$ passes earlier through zero than does i_1, at the point at which i_1 (ignoring sign) is approximately equal to the peak value of i_3. If we represent the phase displacement angle correspondong to this *point* by ψ, then:

$$\sin\psi \approx \psi \approx \frac{I_3}{I} = \frac{1}{9\sqrt{3}} \text{ radians.}$$

We can now recalculate the current through the lamp, bearing this phase displacement in mind, and thus obtain a closer approximation. For this

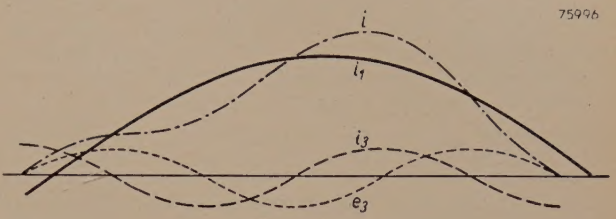


Fig. 9. Combination of the fundamental component i_1 and the third harmonic i_3 to give i, shows that the fundamental current i_1 (which was taken as the lamp current in the first approximation) passes later through zero than i (which serves as a close approximation to the true current: cf. fig. 2). The alternating voltage e_3, which is supplied by the generator with frequency 3ω in the equivalent circuit, passes through zero at practically the same moment as does the total current i. By use of an inductive ballast, i_3 is displaced in phase by approximately 90° with reference to e_3.

purpose, the vector I in fig. 3 must be replaced by I_1, which makes an angle $\psi = 1/9\sqrt{3}$ with V_1, so that $V_s = I_1\,\omega L$ is no longer perpendicular to V_1 (*fig. 10*). Since i_1 passes later than i through zero,

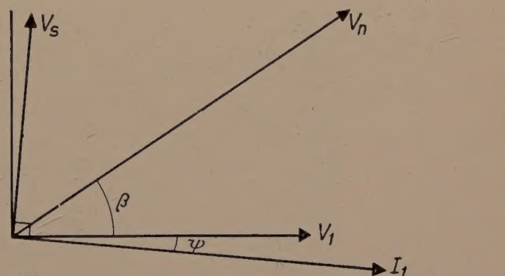


Fig. 10. Vector diagram of the effective values of the currents and voltages in a gas discharge lamp, when the influence of the third harmonic is taken into consideration.

I_1 is slightly behind V_1 in phase. It is therefore more accurate, in this second approximation, not to take a pure resistance $\varrho = aV_0/I$ (formula (3)) as the equivalent element for the lamp, but to place in series with the resistance a small self-inductance λ such that $\omega\lambda/\varrho = 1/9\sqrt{3}$, or:

$$\lambda = 0.06\,\frac{V_0}{\omega I}.$$

Such a series combination of resistance and inductance can, in fact, serve in many cases as a suitable equivalent circuit for a gas discharge lamp, as has been found in practice. It is, however, always only an expedient, in view of the fact that both ϱ and λ depend on the current which has to be passed through the lamp.

Summary. After a short introduction dealing with the properties of gas discharge lamps in general, the characteristic of these lamps is further discussed. If such a lamp is connected to an alternating current supply, the lamp voltage changes, to a first approximation, discontinuously from a positive to a negative value, and vice versa; the lamp voltage is thus a square waveform which can be represented by a *signum* function. As a consequence, higher frequency components, i.e. harmonics, appear in the lamp current. The equivalent circuit of the lamp cannot, for various reasons, including the non-linearity of the characteristic, be composed only of resistances and other linear circuit elements; it can only be represented by a series connection of A.C. generators with frequencies ω (mains frequency), 3ω, 5ω etc. Employing this equivalent circuit, the fundamental component and the third harmonic of the lamp current are calculated to a first approximation. It is found that a gas discharge lamp generates a third harmonic of the order of 6-8% of the fundamental component. The amount of this passing into the mains depends further on the circuit in which the lamp is employed. In conclusion, the current through the lamp is analyzed to a second approximation, from which it is seen that a "substitute" for the equivalent circuit, consisting of a series combination of a resistor and a small self-inductance, can be useful in practice.

A PRE-AMPLIFIER FOR USE WITH ELECTRONIC VOLTMETERS AND OSCILLOSCOPES

by F. G. PEUSCHER and J. van HOLTHOON. 621.394.645:621.317.7

Although the amplifier which is the subject of this article embodies no new principles or revolutionary features, a description would seem to be justified in view of the great utility of this unit, and its value as an adjunct to other electronic measuring equipment.

Triodes and tetrodes were used in valve (or electronic) voltmeters as much as thirty years ago. Instruments of this kind have a number of advantages over other types of A.C. voltmeters. The principal advantages are their high input impedance — resulting in only a small load on the measured voltage — and, in conjunction with amplifiers, their high sensitivity. Especially in the last ten years, the development of electronic voltmeters has made great strides, mainly in the direction of still higher impedances and sensitivities. Of the innumerable types of instrument to which this development has led, the Philips voltmeter type GM 6016 is an outstanding example, being suitable for frequencies of 1 kc/s to 30 Mc/s, and giving full scale deflection on as little as 3 mV.

The advantages mentioned are also relevant to cathode-ray oscilloscopes. These are also used in conjunction with high-impedance amplifiers to give them a high sensitivity, and these amplifiers too, require to be designed for a wide frequency band.

Although sensitivity may already be high, instances occur in practice where even greater sensitivity is required. To meet such needs, an amplifier has been designed [1]) having an input impedance of 5 MΩ and which, connected between a voltage source and a voltmeter, increases sensitivity by a factor of 100. This amplifier (type GM 4574) is depicted in *fig. 1*, and forms the subject of the present article.

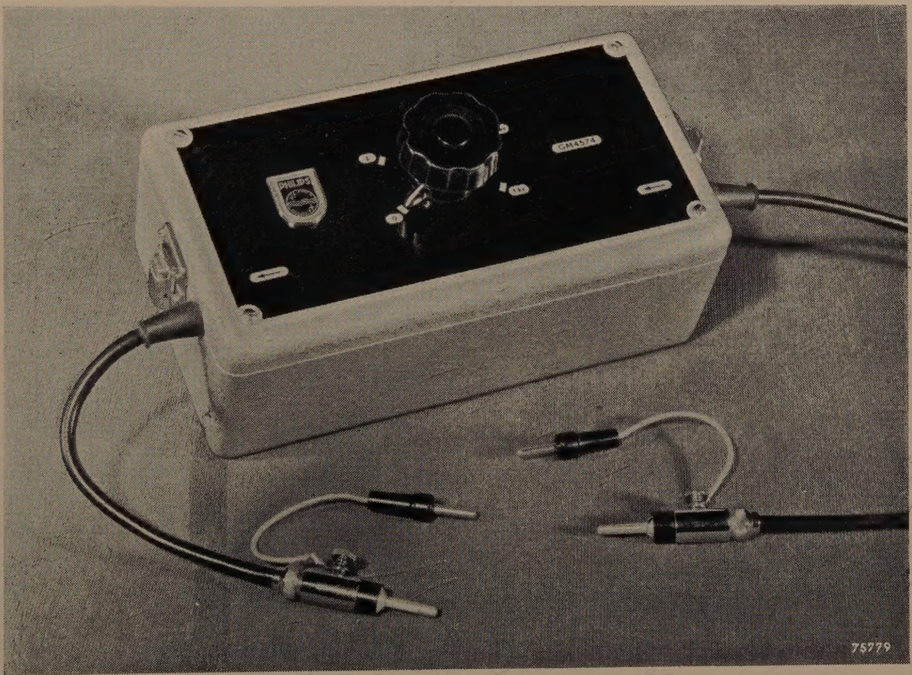


Fig. 1. Amplifier type GM 4574, giving an amplification of 100×. Base 22 cm × 11.5 cm.

[1]) The design was also the work of E. E. Carpentier, who has since left the services of the Company.

General description

In the design of this amplifier, the principal object was to produce a small unit that would not occupy more space than strictly necessary on a bench on which room has also to be found for equipment under test and numerous measuring instruments. For this reason, sub-miniature valves of the kind originally designed for hearing-aids [2]) were decided upon, these being operated from dry batteries. The need for a mains transformer, rectifier valve and smoothing equipment is thus dispensed with. Only a small amount of power is consumed by the filament circuits of these sub-miniature valves, and therefore only small batteries are required.

There is another advantage in the use of batteries: the voltage surges invariably occurring in mains voltages, as well as ripple voltages, are completely avoided. This is very important from the point of view of the purpose for which the amplifier is intended, since the input voltages concerned are often very small, viz. of the order of 0.1 mV.

At the same time, batteries are not without their disadvantages; they tend to produce substances which would set up corrosion or endanger the insulation. In the present amplifier this difficulty has been overcome by housing the batteries in insulated boxes (*fig. 2*) which entirely prevent contamination of the amplifier, and which can be quite easily removed and rinsed in water if necessary.

In order to ensure that the amplification shall remain as constant as possible, this being of course an essential feature where measurements are concerned, considerable negative feed-back is employed. This is particularly desirable because the slope of the valves varies fairly widely as a result of the gradual drop in the battery voltages.

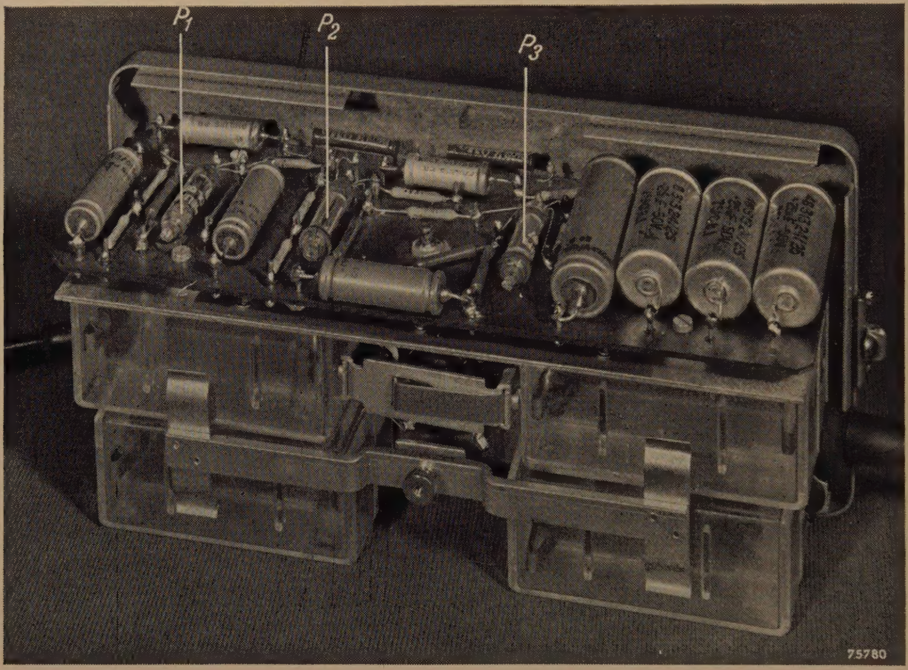


Fig. 2. Interior of the amplifier GM 4574. P_1, P_2, P_3 sub-miniature valves type DL 67. The four boxes made of transparent insulating material are the battery boxes (three for the L.T. batteries (1.5 V), and one for 2 anode batteries of 22.5 V). These boxes ensure effective insulation between the batteries individually and with respect to earth. They also protect the components from contaminating substances from the batteries, and can easily be removed for cleaning.

For an overall amplification of 100×, the gain that would be obtained *without* negative feed-back must accordingly be much more than 100×. Two stages in cascade are provided, each giving an amplification of about 63×, or 4000× in all. These stages precede a third, cathode-follower, circuit, the object of which is to keep the output impedance low (less than 5000 Ω), so that the amplification shall be independent of the (high) input impedance of the various equipment connected to the output terminals. The amplification obtained from this stage is slightly less than unity.

[2]) Philips tech Rev. **15**, 37-48, 1953 (No. 2).

The circuit

The schematic circuit is shown in *fig. 3*. The three valves are pentodes, type DL 67. As the cathodes are not interconnected in this circuit, each filament (current 13 mA) is fed from its own battery (1.5 V). The battery boxes mentioned above guarantee adequate insulation between the batteries themselves and with respect to earth.

The DL 67 is designed to give sufficient anode current even on low supply voltages (max. 45 V), this being ensured by incorporating in it a screen grid of closely spaced wires. The mutual conductance of the screen grid is accordingly very high, which means that variations in screen voltage have a pronounced effect on the working point of the valve. In order to prevent the working point from drifting too much as a result of diminishing voltage from the anode battery, the screen grids of the first and second valves are fed through high resistances (the third valve is connected as a triode). When the battery voltage drops, the screen current also drops slightly and the voltage loss in the feed resistor is reduced, thus partly compensating the fall in the battery voltage. Consequently, the screen-grid current, and hence also the anode current, are much less dependent on the battery voltage than if the screens were connected direct to a tapping on the battery. Differences in the mutual conductance of the screen between one valve and another are thus also smoothed out.

The screens are decoupled by capacitors C_4 and C_5, which are electrolytic capacitors of high capacitance (25 μF), each having in parallel with it a paper capacitor of much lower capacitance (not shown in fig. 3). Electrolytic capacitors have the advantage of a very small volume per μF, especially at the low working voltages occurring in this amplifier. The object of the paper capacitor is to ensure effective decoupling also at high frequencies, at which electrolytic capacitors have a fairly high impedance (their series resistance increases with frequency). A third electrolytic capacitor (C_6) is connected across the anode battery, to avoid undesirable coupling across the internal resistance of the battery, which can become fairly high as the battery gets older.

Capacitive coupling is employed between the input and first valve (C_1-R_1), between the first and second valves (C_2-R_2) and at the output (C_3-R_3); the second valve is D.C.-coupled to the third. The cathode connection is in each case the negative side of the L.T. battery. Grid bias for the first and second valves is established with respect to the negative side of the filament, across the high-value leak resistors (R_1, R_2); this bias is increased slightly for the first valve by the voltage drop across the impedance Z_1 which is part of the negative feed-back circuit.

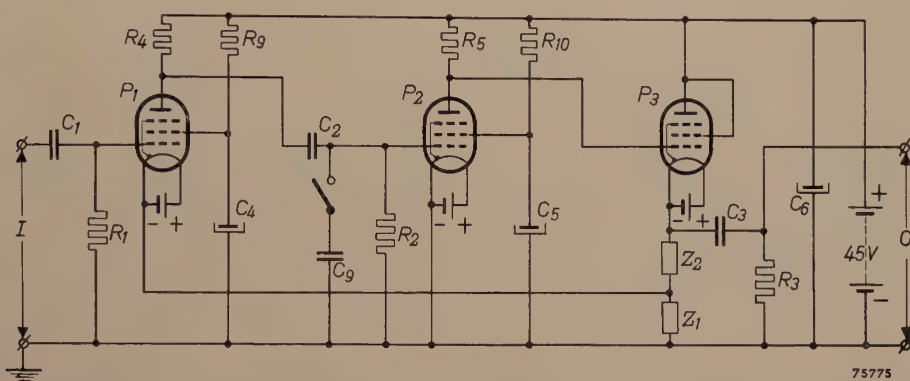


Fig. 3. Circuit diagram of the amplifier. *I* input. *O* output. P_1, P_2, P_3 pentodes type DL 67 (P_3 connected as a triode). C_1, C_2, C_3 coupling capacitors. C_4. C_5, C_6decouplers. C_9 capacitor for suppressing the noise voltage while reducing the band-width. R_1, R_2 grid leaks. R_3 resistor across output terminals. R_9, R_{10} screen grid resistors. Z_1-Z_2 negative feed-back network.

The slope of the valves at the working point is about 0.1 mA/V; the input impedance is roughly 5 MΩ.

Negative feed-back

The negative feed-back circuit is shown in fig. 3. It consists of two impedances Z_1 and Z_2; a feed-back of $\beta = Z_1\ (Z_1 + Z_2)$ is produced, this representing that part of the output voltage which is returned to the input. The ratio in which the amplification A (without negative feed-back) is thereby reduced, is $1{:}(1 + \beta A)$. In the present case A is about 4000 and the required amplification is $A' = 100$, so that $\beta = Z_1\ (Z_1 + Z_2)$ must be roughly 1/100. Variations

in the amplification are thereby reduced in the ratio of $1:(1+\beta A)=1/40$, which is sufficient to ensure that the amplification (except at the very highest frequencies concerned) is not reduced by more than 10% when all the battery voltages have dropped to one half their rated values. Non-linear distortion is also attenuated in a ratio of 1/40, and this is important when the amplifier is employed with an oscilloscope.

Fig. 4 gives further details of the negative feed-back circuit. At frequencies in the region of 1000 c/s the capacitances C_p and C_7 may be neglected, and C_8 can be regarded as a short circuit. Z_1 then consists only of the resistor R_6, and Z_2 of resistors R_7 and R_8 in parallel. The required amount of amplification (100 ± 3%) is obtained by using the appropriate value for R_6.

In fig. 4, C_p represents a stray capacitance, this being mainly the capacitance to earth of the L.T. battery for the first valve. At high frequencies the effect of C_p is not negligible, but this can be counteracted by means of a small trimmer C_7, since the ratio $Z_1/(Z_1+Z_2)$ is independent of frequency when C_7 satisfies the condition

$$C_7 \frac{R_7R_8}{R_7+R_8} = C_pR_6.$$

To ensure that voltages of square waveform will be amplified without distortion, C_7 is very carefully adjusted, using a square-wave voltage of about 5000 c/s. The form of the output voltage of the amplifier is simultaneously checked with an oscilloscope which immediately reveals any discrepancy in the adjustment [3]) (see *fig. 5*).

network is so designed that the feed-back is reduced with decreasing frequency. When the frequency is reduced, the capacitor C_8 (fig. 4) increases the impedance of the branch R_8-C_8 and hence also of Z_2, so that the ratio of $Z_1/(Z_1+Z_2)$ decreases.

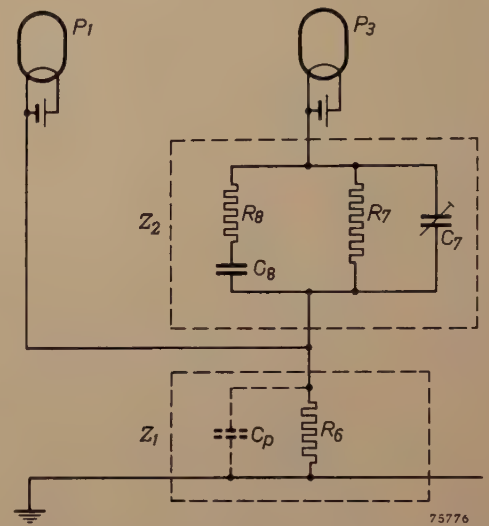


Fig. 4. Details of the negative feed-back network Z_1-Z_2. C_p is a stray capacitance in parallel with R_6; the effects of this capacitance are counteracted by means of a trimmer C_7. The fixed capacitor C_8 in series with R_8 improves the response curve at low frequencies.

This compensation is not perfect, however; to ensure that there will not be too little compensation in any given frequency range, some over-compensation must be permitted in another range. This is the reason why the curve (*1* in *fig. 6*) rises slightly at frequencies below 100 c/s, to reach a maximum

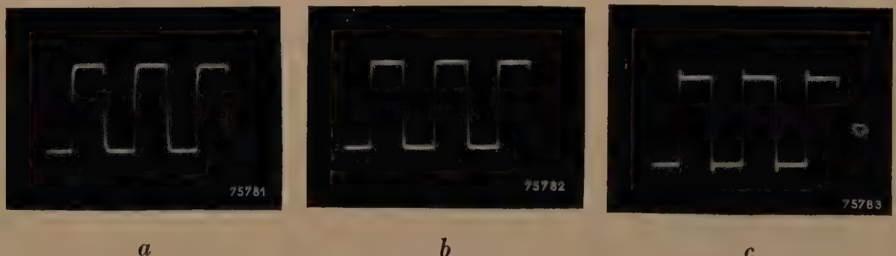


Fig. 5. Effect of the trimmer C_7 (fig. 4). The input voltage is of square waveform, 5000 c/s. Oscillograms of the output voltage: *a*) value of C_7 too low, *b*) correct, *c*) too high. Waveforms displayed on oscilloscope type GM 5653.

Owing to the finite time constant of the three *RC* couplings (C_1-R_1, C_2-R_2 and C_3-R_3), the gain is smaller in the lower frequencies than in the middle of the range; to compensate this, the feed-back

at 2 c/s, which on average over a number of these amplifiers is roughly 12% above the nominal value of the amplification. Consequently, a perfectly square wave of 12.5 c/s is slightly distorted, as will be seen from *fig. 7a*; at 25 c/s this distortion is almost imperceptible, and at 50 c/s it cannot be

[3]) See J. Haantjes, Judging an amplifier by means of the transient characteristic, Philips tech. Rev. 6, 193-201, 1941.

detected at all (fig. 7*b* and *c*). Without the compensation, even a 50-cycle square-wave voltage would be badly distorted, and the compensation therefore ensures a useful increase in the range of frequencies for which the amplifier can be employed.

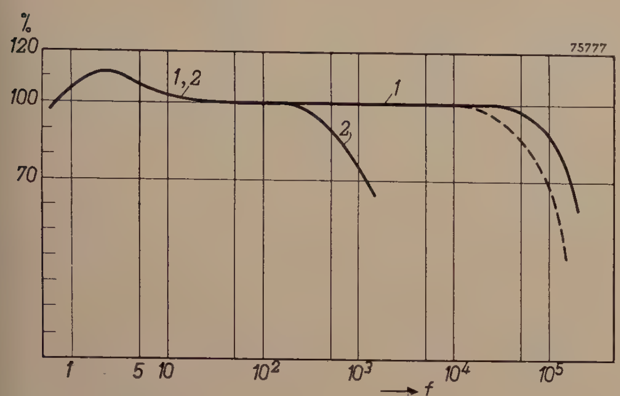


Fig. 6. Amplification (relative) as a function of the frequency, plotted for a load of 1 MΩ in parallel with 20 pF: the curve *1* is obtained without the capacitor C_9 in circuit, *2* is obtained with C_9 connected (fig. 3). 100% corresponds to amplification $A' = 100 \pm 3\%$, with nominal battery voltages, and with 10% less than this value for battery voltages of half the rated values. In the latter case the curve is shown by the broken line.

With a response curve of this shape (*1* in fig. 6), obtained by means of the compensation mentioned, the unit is suitable for amplifiying sinusoidal frequencies of from 1 to 150,000 c/s; with a square wave voltage the frequency range is 10 to 10,000 c/s.

Even when all the battery voltages have dropped to half their rated values, the response curve retains the form shown, with the exception of the extreme right-hand end, i.e., at the higher frequencies, where the drop becomes slightly steeper (broken line in fig. 6).

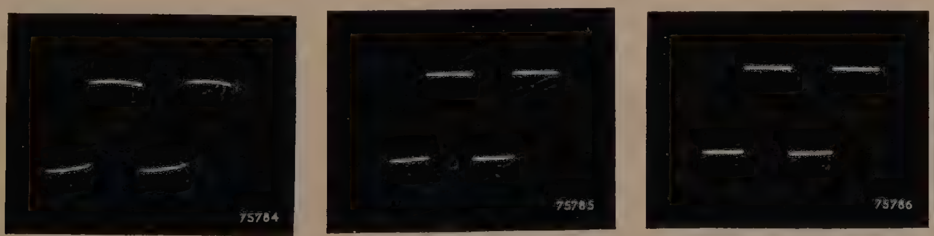
frequencies; the high-frequency end of the range can then be safely suppressed, thus eliminating the noise attributable to the higher frequencies. This is effected very simply by placing a capacitor in circuit (C_9 in fig. 3). The noise voltage, referred to the input, is less than 10 μV without the capacitor C_9, but is less than 5 μV with this capacitor in circuit. In this case the response curve of the amplifier is as shown by line *2* in fig. 6, this being suitable for the amplification of sinusoidal voltages at frequencies of 1 to 1000 c/s, and square-wave or pulse voltages of from 10 to 50 c/s.

In order to avoid possible trouble due to microphony when very low voltages are to be amplified, the amplifier chassis is flexible mounted in the case.

"Forming" of the electrolytic capacitors

As already mentioned, the screen grids of the valves are fed through high-value resistors with decoupling by means of electrolytic capacitors (C_4 and C_5, fig. 3). This arrangement necessitates certain precautions in view of the leak current of the capacitors, but which are fully justified by the advantage of electrolytic capacitors mentioned above (small volume per μF).

The dielectric in an electrolytic capacitor is a thin film of aluminium oxide on the aluminium anode [4]), this being produced by a "forming" process which consists in passing a direct current through the capacitor in the correct direction; oxygen is then evolved at the anode, and this oxidizes the aluminium. The higher the D.C. voltage across the capacitor, the thicker the layer of oxide and, hence, the



a *b* *c*

Fig. 7. Oscillograms of the output voltage as obtained from a square waveform voltage, *a*) of 12.5 c/s, *b*) of 25 c/s, *c*) of 50 c/s. At 12.5 c/s there is slight distortion (due to overcompensation by C_8 (fig. 4)); at 25 c/s this has almost disappeared, and at 50 c/s it is imperceptible. Taken with oscilloscope type GM 5653.

Where amplification of very small voltages within a limited range of frequencies is required, it is useful to be able to reduce the background (noise) at the expense of the bandwidth. Such conditions will often occur in investigations of mechanical phenomena, which are concerned with very low

higher the resistance which it offers and the lower the capacitance.

During such time that the capacitors are not in use, the oxide layer is partially broken down, so that,

[4]) W. Ch. van Geel and A. Claassen, Electrolytic condensers, Philips tech. Rev. 2, 65-71, 1937.

when a voltage is applied across these capacitors after a long period of idleness, the leak current is at first quite high. A new layer is quickly formed, however, of thickness depending on the applied voltage.

It follows, therefore, that there may be some difficulty when an amplifier designed on the lines of the circuit shown in fig. 3 is switched on. A leak current is then able to flow through C_4 and C_5 which is of the same order as the screen-grid current (normally 10 μA), and the required screen voltage is thus by no means reached. Moreover, with the potential of only a few volts then occurring across the capacitor, there is little opportunity for the leak current to decrease within a short time owing to re-formation of the capacitor, and thus to establish the required working conditions.

This difficulty is overcome by including two intermediate positions in the switch, which is in any case needed for the filament circuits and for the capacitor C_9. The switch accordingly has five positions, which we shall now mention in turn.

In the first position ("0"), the three filament circuits are open and the anode battery is switched off, this being necessary to prevent discharge due to the leak currents in the three electrolytic capacitors when the amplifier is not in use.

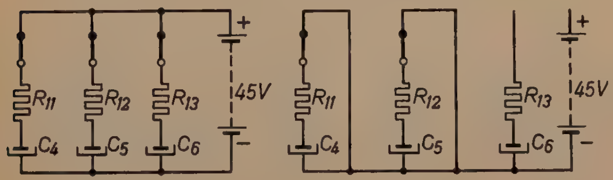


Fig. 8. *a*) In position I of the switch, the electrolytic capacitors C_4, C_5 and C_6 are "formed" at 45 V, through resistors R_{11}, R_{12} and R_{13}.
b) In position II, C_4 and C_5 are discharged through R_{11} and R_{12}.

In the second position (I) the circuit depicted in *fig. 8a* is established in order to form the electrolytic capacitors at 45 V through current-limiting resistors (C_6 is included in this to save the battery from the high initial leak current that would flow through this capacitor if placed in circuit at once). In a short time (a few seconds if the amplifier has not been used for some weeks, or about 1 minute if idle for longer periods), the leak currents are so small that the capacitors are charged almost to the voltage of the anode battery.

If the switch were now set to the normal working conditions (fig. 3), an initial voltage of 45 V would be applied to the screen grids of the first and second valves, by reason of the charge on C_4 and C_5, and this is much higher than the normal potential, which is only 10 V. The resultant heavy screen current would be detrimental to the valves, and the switch is therefore so arranged that in the third position (II), the capacitors C_4 and C_5 are discharged through resistors (fig. 8*b*).

These capacitors are connected to the screen grids only in the fourth position ("100 kc/s"), with C_6 directly across the anode battery. At the same time all the filament circuits are closed, giving the arrangement as shown in fig. 3.

The amplifier then functions in accordance with the response curve *1* in fig. 6. Once the capacitors C_4 and C_5 have been re-formed at 45 V with the switch in position I, the leak current at the 10 V now applied to them is quite low. In the long run, however, at this low working voltage, the layer of aluminium oxide becomes thinner and the leak current accordingly higher, and it is therefore advisable when the amplifier is continuously in use, to turn the switch to position I for a few moments, from time to time, to re-form C_4 and C_5 at 45 V. With continued use, a few minutes every fortnight is sufficient for this purpose.

In the fifth position of the switch ("1 kc/s"), C_9 is in circuit (see fig. 3), the response curve is in accordance with curve *2* in fig. 6 and the noise level is reduced.

Summary. An amplifier (type GM 4574) is described, designed for increasing the sensitivity of electronic voltmeters and oscilloscopes. Sub-miniature valves operated from dry batteries ensure a light, compact unit. The amplifier comprises two initial stages each giving amplification of about 63×, and a low impedance cathode-follower output stage (less than 5000 Ω). The use of considerable negative feed-back reduces the overall amplification in the straight part of the characteristic to 100 ± 3%, but renders the amplification insensitive to variations in the valve characteristics and battery voltages (amplification is reduced by at most 10% by a drop in battery voltage to half the rated value), and reduces non-linear distortion. A further object of the negative feed-back circuit is to correct the response curve. The amplification as a function of the frequency is illustrated; it is roughly 100× throughout the greater part of the range of from 1 to 150,000 c/s, with a maximum of 112× at 2 c/s and a minimum of 75× at 150,000 c/s. Noise level can be reduced by narrowing the frequency range to 1-1000 c/s. For the amplification of square-wave and pulse voltages the wide frequency band is 10-10.000 c/s and the narrow band 10-50 c/s. The noise voltage, referred to the input, is less than 10 μV in the wide band and less than 5 μV in the smaller one.

THE TROPOSPHERE AS A MEDIUM FOR THE PROPAGATION OF RADIO WAVES - II

by H. BREMMER.

621.396.11:551.510.52

In the first part of this article [13]*) six separate mechanisms were distinguished by which radio communication may be effected by microwaves* *)*. Five of them were discussed in the first article, and the present article is wholly devoted to the remaining one, namely scattering. Such a detailed treatment is justified, not because scattering is important in itself for radio communication — it is too uncertain — but because it is largely responsible for interference in the microwave region from distant transmitters. It is therefore extremely important for such practical problems as allocating wavelengths among existing or future transmitters, and determining the minimum separation of two transmitters of the same wavelength if mutual interference is to be avoided.*

Scattering by inhomogeneities in the troposphere

So far a number of idealized cases have been distinguished whereby microwaves can be propagated through the troposphere. In each of them the refractive index depends only upon the height above the earth's surface. This was discussed in detail in the first part of this article [13]) (hereafter referred to as I). It was stated that under normal conditions, a field beyond the horizon to the transmitter was produced mainly by diffraction. However, fields have repeatedly been observed beyond the horizon of the transmitter that are much too strong to be explained by diffraction. The way in which these fields vary suggests that they are caused by inhomogeneities in the troposphere. There are always chance irregularities produced by solids, such as aeroplanes, birds, insects etc, but these cannot explain the systematically fluctuating fields, that can interfere with television reception, for example, up to several hundred kilometres from a microwave transmitter.

Fields like this may be generally attributed to scattering, i.e. to local variations in the refractive index due to small-scale inhomogeneities in the troposphere. Their scattering action on light is to be seen in the twinkling of stars, which seems to originate mainly in the lowest layer of the troposphere. Similar scattering effects are experienced by radio waves. According to a theory of Booker and Gordon [14]), it seems that the order of magnitude of the fields due to scattering agrees with the (rather scanty) meteorological data on the inhomogeneities. These data are more especially concerned with fluctuations in the wind speed, the temperature and the humidity in the lowest layer [15]).

For scattering of radio waves, only the fluctuations δn in the refractive index are important, and δn, according to formula (1) in I, results from separate fluctuations δp_a of the pressure of the dry air, δp_w of the water vapour pressure and δT of the temperature. For a typical situation with $T = 300\ ^\circ\mathrm{K}$, $p_a = 1000$ millibars and $p_w = 20$ millibars, we find:

$$\delta n = (0.26\ \delta p_a + 4.2\ \delta p_w - 1.44\ \delta T)\ 10^{-6}. \quad . \quad (7)$$

Temperature and pressure change so rapidly that heat-exchange with the surroundings can be neglected. On the average, therefore, relationships hold good between the fluctuations δp_a and δT, and also between δp_w and δT, that agree with the relationship between pressure and temperature for adiabatic changes. With the help of these, (7) can be simplified to

$$\delta n \approx -2\ \delta T \times 10^{-6}. \quad . \quad . \quad . \quad . \quad . \quad (8)$$

Thus readings of temperature fluctuations give, in themselves, an idea of what fluctuations to expect in n. Gerhardt and Gordon [16]) have made observations in Texas of the variation of temperature in a small arbitrary region of the atmosphere under normal conditions. They observed temperature fluctuations of the order of 1 °C, with time constants (depending on weather conditions) between a few minutes and roughly a second. According to (8), these measurements correspond to fluctuations in n of the order of 10^{-6}, which is several hundred

*) Note that "microwave" is again used in an extended sense, to embrace all waves of $\lambda < 10$ metres.

[13]) This Review, November 1953, p. 148. Note that on p. 151 of this article (I), eight lines below fig. 3, a printing error occurs: for $\lambda = 5$ cm, read $\lambda = 5$ m.

[14]) H. G. Booker and W. E. Gordon, Proc. Inst. Rad. Engrs. **38**, 401-412, 1950.

[15]) A review is to be found in e.g. O. G. Sutton, Atmospheric turbulence, Methuen, London 1949, Chapter 2.

[16]) J. R. Gerhardt and W. E. Gordon, J. Meteor. **5**, 197-203, 1948.

times smaller than the normal value of $n-1$ (about 0.0003).

More recently Birnbaum [17]) has measured the fluctuations in n directly from the difference between the resonance frequencies of two cavity resonators, one sealed and the other open to the atmosphere. In this way he measured the maximum distance for which a correlation still existed between the fluctuations at two separate points of measurement. This distance varied, depending on the type of weather, between values of the order of 10 metres and 10 centimetres. It fixes the dimensions of what we can term separate "eddies", and the relation between these dimensions and the wavelength is all-important for the scattering phenomena. We shall come back to this point, but we shall first consider the physical conditions that determine the sizes of eddies.

Inhomogeneities caused by turbulence

The eddies are in turbulent motion with respect to each other. The eddies themselves may be of any size, from very great (e.g. atmospheric depressions) to very small, of the order of centimetres. It is the small ones that particularly interest us [18]).

In an extended medium like the atmosphere, turbulent movements can apparently be stable, provided energy is continually brought to them from outside (in the present case, radiation from the sun). When a statistical equilibrium is attained, this energy is finally dissipated as heat, through friction due to the turbulence. The dissipation is greatest for the smallest eddies, for these are responsible for the largest velocity gradients. A rapid transport of energy is therefore conducive to the formation of *small* eddies, for only then can enough friction develop to dissipate all the energy. As to the influence of viscosity, friction opposes the formation of very small eddies, since it slows down the consequent strong turbulent movements. It follows that the size l_η of the smallest eddies that play a significant rôle in the energy exchange process decreases when the rate of arrival of energy becomes greater, and increases with increasing viscosity. This agrees with the following formula, derived by Taylor [19]), which relates the "eddy size" l_η (defined by him more precisely on a statistical basis) to the coefficient of viscosity η, the energy ε absorbed per second per unit volume, and the mean kinetic energy $\frac{1}{2}v^2$ per unit mass of turbulent gas:

$$l_\eta = \sqrt{5\,\frac{\eta}{\varepsilon}\,v^2}.$$

The phenomenological theory leading to this result assumes an extended medium, large compared with the average size of an eddy, and further, that the turbulence is isotropic and that the effect of compressibility of the medium may be neglected (the latter is of importance only for air speeds comparable with the speed of sound). However, this theory says nothing about the statistical size-distribution of the eddies, and to do this, further physical assumptions have to be made. Using only simple dimensional arguments, Kolmogoroff arrived at a distribution for which the probability of obtaining a small eddy of size l (within which eddy the turbulence speed must be about the same everywhere) is proportional to $l^{5/3}$, but he neglected the influence of friction. After allowing for the friction, Heisenberg found that the probability for very small eddies decreased far more rapidly, namely in proportion to l^7, and that the rapid decrease begins with eddy sizes of the order of $l_{\text{crit}} = (l_\eta)^{3/2}/L^{1/2}$, where L is the maximum dimension in the medium. Friction thus imposes a limit not only on the size l_η of eddies that take a significant part in energy-exchange ("macro-eddies"), but also on the smallest eddy size l_{crit} that has an appreciable probability of existing ("micro-eddies"). The numerical value of l_{crit} is very important for scattering phenomena, as we shall see later.

Before going on to the calculation of fields due to scattering, we quote a verse attributed to L.F. Richardson, which is a variant on a well known jingle [20]). It pithily sums up what happens to the energy taken up by a turbulent medium:

Big whirls have little whirls
That feed on their velocity;
Little whirls have lesser whirls,
And so on to viscosity.

Determination of the scattering field due to a single volume element

We assume that the refractive index varies slowly enough with time for it to be regarded as constant during one period of the transmitter frequency, and further that a perceptible correlation between the values of δn at two points exists only for points

[17]) G. Birnbaum, Phys. Rev. **83**, 110-111, 1951.
[18]) A very important review of turbulence phenomena is to be found in an article by S. Chandrasekhar, Astrophys. J. **110**, 329-339, 1949.
[19]) G. I. Taylor, Proc. Roy. Soc. **A 151**, 421-444, 1935 (Equation 50).

[20]) Big fleas have little fleas
Upon their backs to bite 'em;
Little fleas have lesser fleas,
And so — ad infinitum.

within a certain distance of each other — of the order of magnitude of l_{crit}.

The contributions to the field originating from scattering of the primary beam by points P within a volume element dV not smaller than a micro-eddy, may be assumed to be coherent. In contrast to this, the separate net contributions from these volume elements are to be regarded as mutually incoherent. (It should be noted that as rapid fluctuations are involved, only statistical averages of these contributions have any meaning.) As a result of this incoherence, the amplitudes of the contributions dE to the field H should not be summed but their squares, that is, their intensities.

The first step of the calculation therefore requires the intensity of the scatter field caused by a single volume element dV. The calculation is greatly simplified by assuming that the dimensions of dV are small compared with the distances TP and PR from the volume element to the transmitter and to the receiver, respectively (*fig. 8*). Just as

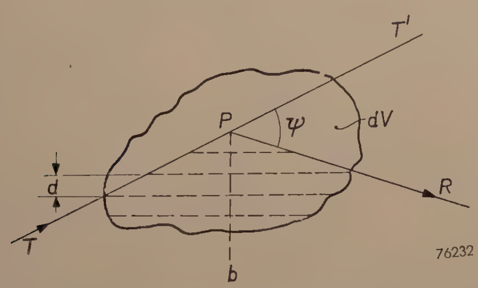


Fig. 8. The volume element dV represents a scattering element in the troposphere. TP incident ray, PR scattered ray. Angle $T'PR = \psi$ is the scattering angle. The dashed lines indicate a few planes at right angles to the bisector b of the angle TPR. The mutual separation d of these planes satisfies Bragg's law (eq. 9).

with X-ray scattering in crystals, the scattered intensity as a function of direction (RP) depends only on the distribution in space of δn which produces the scattering, where δn is, of course, a function of (x, y, z). More precisely, the intensity at R depends on only one component of the (here continuous) Fourier spectrum of δn within dV with respect to the spatial co-ordinates x, y and z. (The value of this component will be given later.) This means that a fictitious δn distribution with only this one Fourier component would produce the same scattering as the real δn distribution (in so far as this approximation is a good one). This component now corresponds to a distribution for which equal values of δn occur in planes (indicated by dotted lines in fig. 8) at right angles to the bisector b of the angle TPR: their mutual separation d is given by

$$d = \lambda/(2 \sin \tfrac{1}{2} \psi). \quad . \quad . \quad . \quad . \quad . \quad (9)$$

The scattering angle ψ is here the supplement of the angle TPR.

We recognize (9) as Bragg's law. It means that the Fourier component that completely determines the scattering corresponds to a certain array of planes with periodic values of δn, and these surfaces cause, by interference, a maximum intensity in precisely the direction PR. The analogy with Bragg's law is only possible because the fluctuations can be regarded as changing relatively slowly with time, such that the distribution of δn during a number of consecutive periods of the transmitter frequency can be thought of as fixed. The scattering is then as in a solid. An important difference from scattering by crystals is, however, that there is no orderly cell structure, so that the intensity of the scattered waves changes continuously with the direction of observation. In this respect it is more like scattering in an amorphous solid.

It follows that the intensity d(E^2) of the contribution to the field from a single elemental volume dV is in the first place proportional to the intensity of the relevant Fourier component. The magnitude of d(E^2) therefore depends on the location of the transmitter and the receiver, since the position of the Fourier component in the spectrum changes with these locations.

The theory also tells us that d(E^2) also depends upon the physical factor $(\overline{\delta n/n})^2$, which is the mean-square of the relative deviation of the refractive index. This makes it possible to draw a comparison between scattering of radio and of sound waves. For radio waves, according to (8), $\delta n/n$ is equal to $-2 \times 10^{-6}\, \delta T$, whilst for sound waves, for which n is proportional to $T^{-\frac{1}{2}}$, $\delta n/n$ is equal to $-\frac{1}{2}\delta T/T$. Under like circumstances, that is to say, the same condition of the troposphere and the same wavelength (it so happens that sound waves have lengths varying from a few centimetres to a few metres also), sound wave scattering produces a field intensity per unit volume greater by a factor $10^6/4T$, which is of the order of $1000\times$.

The total scattering-field at a receiver

The incoherent contributions of the separate volume elements lying between the transmitter and the receiver have to be summed by integrating d(E^2) over the whole scattering space. We make the mathematics less difficult by confining ourselves to the case of a transmitter T beamed on a receiver

R (*fig. 9*). We shall assume that both have polar diagrams with half-value angles [21]) of $\Theta_{\frac{1}{2},\mathrm{h}}$ in the horizontal plane, each bisected by the vertical plane through T and R, and a half-value angle $\Theta_{\frac{1}{2},\mathrm{v}}$ in the vertical plane with respect to the local horizon. It is assumed that $\Theta_{\frac{1}{2},\mathrm{h}}$ and $\Theta_{\frac{1}{2},\mathrm{v}}$ are both small angles. As far as reception is concerned, only

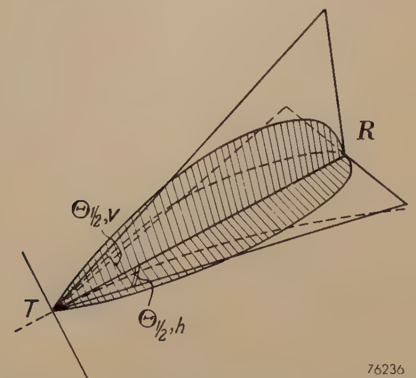


Fig. 9. Transmitter T with beam directed at the receiver. The beam has a polar diagram in the horizontal plane with half-value angle $\Theta_{\frac{1}{2},\mathrm{h}}$ which is bisected by the vertical plane through transmitter and receiver, and a polar diagram in the vertical plane with half-value angle $\Theta_{\frac{1}{2},\mathrm{v}}$ with respect to the horizontal planes. The angles $\Theta_{\frac{1}{2},\mathrm{h}}$ and $\Theta_{\frac{1}{2},\mathrm{v}}$ have been exaggerated in the diagram for clearness. The receiver is assumed to have a similar polar diagram.

scattering occurring within the volume defined by the intersection of the transmitter and receiver half-value angles is of importance. In *fig. 10* the cross-section of this volume in the plane through T and R — the plane of the diagram — is shown shaded.

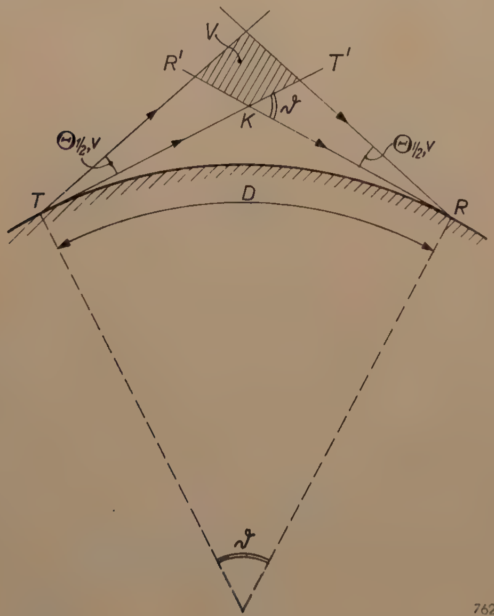


Fig. 10. From the point of view of reception, for beams like those in fig. 9, the scattering is important only when it takes place within a volume V; its cross section in the plane of the paper is shaded. Angle $T'KR$ = angle $TOR = \vartheta$.

[21]) The half-value angle is the angle between directions for which the intensity is half that along TR (fig. 9).

Booker and Gordon [14]) have treated many cases, of which we shall only consider those for great distances, for then the scattering fields are relatively the most important. The scattering volume is then confined to the neighbourhood of the plane bisecting TR, and it is small compared with the distance $\frac{1}{2}D = \frac{1}{2}a\vartheta$ to transmitter and receiver (D = arc TR, ϑ = angle TOR, a = radius of the earth), provided that

$$\vartheta \gg \Theta_{\frac{1}{2},\mathrm{v}} \text{ and } \Theta_{\frac{1}{2},\mathrm{h}} \ll \tfrac{1}{2}\pi .$$

All the scattering angles are then approximately the same, and we can take them equal to $T'KR$, (fig. 10), corresponding to the rays TK and KR, which are horizontal at the transmitter and receiver respectively. As is clear from the figure, the angle ψ (of fig. 8) is equal to the angle ϑ subtended by transmitter and receiver at the earth's centre. Since the field intensity depends on D, the total scattering field depends on the angle ψ.

This raises the question: for what values of ψ can appreciable scattering be expected? We can answer this qualitatively, provided that δn changes only very slightly over distances of the order of l_{crit}. When this is so, the components of the volume Fourier spectrum of δn must of course be negligible for periods less than l. Using this in eq. (9), when only Fourier components with periods greater than l contribute to the scattering, the corresponding angles ψ must satisfy

$$\frac{\lambda}{2\sin\frac{1}{2}\psi} > l, \text{ or } \sin\tfrac{1}{2}\psi < \frac{\lambda}{2l}. \quad . \quad . \quad (10)$$

The condition is easiest to see from its two limiting cases:

1) $\lambda \gg l$. The scattering is appreciable in all directions satisfying (10). This is the well-known Rayleigh scattering, which is more or less isotropic.
2) $\lambda \ll l$. Scattering of importance occurs only through very small angles ψ; the scattering therefore takes place mainly in the forward direction.

For great distances with a large angle ϑ, the most marked contribution to the field at the receiver is found in the first case ($\lambda \gg l$), where there is no limit on the scattering angle. The importance of the parameter l now becomes evident; this is apparently of the order of l_{crit}, that is, the smallest eddy size that is at all probable (micro-eddies, see the end of the section on turbulence).

Evaluation of the scattered field

We have just seen that the total scattered field depends, in particular, upon one Fourier component of the distribution in space of δn in a coherently scattering volume δV. It appears from a fuller investigation that this dependence can be explained by statistical properties of the space distribution of δn. All we need to know is the following so-called correlation function

$$C(a,b,c)=\frac{\overline{\delta n(x,y,z)\times\delta n(x+a,\ y+b,\ z+c)}}{\overline{\{\delta n\ (x,y,z)\}^2}}.$$

In this definition the bars denote means over the volume elements, which must not be smaller than a micro-eddy. This function gives the correlation between the fluctuations of n at two points with cartesian co-ordinates differing by a, b and c respectively, so that for isotropic turbulence they depend only on the distance $(a^2+b^2+c^2)^{\frac{1}{2}}$. At very large distances there is no correlation and the correlation function decreases to zero. The function is also normalized; its maximum value is at zero distance and is there unity. Apart from these limiting values, the function decreases monotonically with increasing distance. As examples of mathematically tractable cases we mention two isotropic ones

$$C(a,b,c)=\mathrm{e}^{-(a^2+b^2+c^2)^{\frac{1}{2}}/l} \quad \ldots \quad (11a)$$

and

$$C(a,b,c)=\mathrm{e}^{-(a^2+b^2+c^2)/l^2} \quad \ldots \quad (11b)$$

In both of them the correlation begins to decrease sharply when the distance $(a^2+b^2+c^2)^{\frac{1}{2}}$ becomes of the order of l, so that l can be interpreted as a measure of the average micro-eddy size. The field strengths E corresponding to these correlation functions at great distances from the transmitter can now be found. For a transmitter radiating a power of 1 kW, these field strengths are respectively

$$E=\frac{4.2\times10^8\times l_{\mathrm{cm}}^{\frac{3}{2}}\left\{\overline{(\delta n)^2}\,\Theta_{\frac{1}{2},\mathrm{v}}\right\}^{\frac{1}{2}}}{\lambda^2_{\mathrm{cm}}\,D_{\mathrm{km}}\left\{1+\dfrac{16\pi^2l^2}{\lambda^2}\sin^2\left(\dfrac{D}{2a}\right)\right\}}\ \mathrm{mV/m}, \qquad (12a)$$

$$E=\frac{2.0\times10^8\times l_{\mathrm{cm}}^{\frac{3}{2}}\left\{\overline{(\delta n)^2}\,\Theta_{\frac{1}{2},\mathrm{v}}\right\}^{\frac{1}{2}}}{\lambda^2_{\mathrm{cm}}}\times\frac{\mathrm{e}^{-\frac{2\pi^2l^2}{\lambda^2}\sin^2\left(\frac{D}{2a}\right)}}{D_{\mathrm{km}}}\ \mathrm{mV/m}. \quad \ldots \quad (12b)$$

(E is the r.m.s. field strength in millivolts per metre. Units of length are given as suffixes, except in dimensionless ratios of lengths.) These formulae clearly demonstrate the influence of the form of the correlation function. If $\lambda \ll l$, the first case corresponds to a decreasing field approximately proportional to D^{-3}, and the second case to a field that decreases roughly in proportion to $D^{-1}\exp(-\gamma D^2)$, where γ is a positive constant.

Interpretation of observations in terms of scattering

Although so much depends on the precise form of the correlation function C, which is not known, field strengths far beyond the horizon in general appear to be satisfactorily explained by the Booker-Gordon theory [14]). This theory starts from the correlation function (11a) and leads to a field like that in (12a) that decreases not exponentially, as does a diffraction field, but more slowly.

The theoretical results can also be compared with observations relating to the distance up to which television reception from one transmitter can be interfered with by another working at the same wavelength. As with the analysis of ionospheric propagation data, from averages of a large number of observations, field strength - distance curves can be plotted, which give the field strengths exceeded during e.g. 50%, 20% or 10% of the time. For any one distance, the greater the field strength, the smaller the percentage and the smaller the influence of the diffraction field. The latter varies as the effective radius of the earth gradually changes, but these changes are certainly not large enough to explain the rather large field strengths that are observed only during, say, 1% of the time. It appears that these field strengths are to be attributed entirely to sporadic, sometimes very strong, scattering phenomena.

In this connection the curves obtained by the Comité Consultatif International des Radiocommunications [22]) may be mentioned. These are for fields from a half-wave dipole, exceeded during 1% and during 10% of the time. For example, for distances D between 100 and 500 km, and frequencies between 50 and 200 Mc/s, it appears that the field strength E exceeded during 1% of the time can be fairly accurately represented by

$$E=85\,000\ D^{-3}\ \mathrm{mV/m}$$

(with D in km, and again for 1 kW radiation). The factor D^{-3} is consistent with (12a) provided $l \gg \lambda$, for then the term 1 in the denominator can be neglected, and also $\sin D/2a$ can be replaced by $D/2a$, so that a factor D^3 appears in the denominator. However, this does not confirm the fundamental assumptions underlying (12a), for a similar decrease of field strength can be obtained, for example, if the turbulence is not isotropic.

[22]) C.C.I.R. Documents of the 6th plenary assembly, Geneva 1951, page 55.

This is an important point. E. G. Richardson [23]) has made direct measurements of air turbulence in the lowest layers, and these show not only that the eddy size increases with height, but also that its isotropic character, in so far as it exists close to the earth's surface, is lost. Moreover, from measurements in aircraft (N. S. Wong), it now seems likely that the refractive index varies far less horizontally than vertically. This is also confirmed by Herbstreit and Norton, who studied the correlation between the field strengths from a transmitter, measured at two receivers placed close to each other. The correlations were obtained for different directions of the line joining the receivers, and comparing them it appears that the correlation extends over a greater distance horizontally than vertically.

These observations led Feinstein to reconsider the simple model of a refractive index depending only on the height, and to assume in addition that the vertical gradient of the refractive index fluctuates around a constant mean. The theory thus assumes a strongly anisotropic turbulence. With suitable assumptions concerning the ratio of the mean vertical dimension of the eddies to the wavelength, this theory, too, can yield a field strength proportional to a negative power of the distance D, as does the theory based on isotropic turbulence. Field strength measurements at large distances are therefore still insufficient to decide between isotropy and anisotropy. More comprehensive observations on scattered fields, combined with meteorological data, are required to extend our knowledge of turbulence.

Fading explained by scattering

The scattering mechanism can also explain fading, which occurs especially at great distances. In fact, scattering apart, fading is always liable to occur whenever field strengths of the same order of magnitude are contributed by two or more different modes of propagation (e.g. direct paths, reflections at transition layers, diffraction, etc). These contributions depend on nearly independent physical conditions, and this can make their phase differences change very quickly, so that the total field strength fluctuates violently. However, for great distances, where scattering effects dominate all other contributions to the field, fading can be explained successfully only by scattering, according to the mechanism first proposed by Ratcliffe [24]) for ionospheric fading.

[23]) E. G. Richardson, Proc. Roy. Soc. A **203**, 149-164, 1950.
[24]) J. A. Ratcliffe, Nature **162**, 9-11, 1948.

The eddies in the troposphere can be regarded as stationary during one period of the high-frequency radiation, but in fact they do move slightly. Even when their movement is so slow that it does not influence the average field, the scattered field due to the eddies will be subjected to a Doppler effect, and even this can cause small variations of field strength about the mean value. If v_b is the component of the velocity in the direction of the bisector of the incident and scattered rays (direction b in fig. 8), and c is the velocity of light, it is found that for a scattering angle ψ, a Doppler displacement $\Delta\nu$ with respect to the frequency ν of the carrier wave exists, such that

$$\frac{\Delta\nu}{\nu} = \frac{2v_b}{c}\sin\tfrac{1}{2}\psi .$$

The field is thus modulated with a frequency $\Delta\nu$, and fading develops with a time constant

$$T = \frac{1}{\Delta\nu} = \frac{\lambda}{2v_b \sin\tfrac{1}{2}\psi} .$$

For large distances ($\vartheta \gg \Theta_{\frac{1}{2},\mathrm{v}}$), ψ can be put equal to the angle $\vartheta = D/a$ between transmitter and receiver. Fading then occurs with a time constant dependent on the tropospheric parameter v_b, and on the value of the fixed parameter λ/D. The tempo of the fading fluctuations thus becomes quicker with increasing distances. An analysis of the fading phenomena could therefore provide information on the air velocities characterising the turbulence.

Recapitulation

In conclusion, this section recapitulates the many different factors that control the propagation of short radio waves in the troposphere. Their distinguishing characteristics are outlined.

Communication between a transmitter and receiver comes about in one of the following ways, or through a combination of them:

a) *Direct propagation:* the total field is due to two rays, both being refracted according to Snel's law: one is the direct ray and the other undergoes a reflection at the earth's surface. *Occurrence:* whenever the receiver is above the horizon of the transmitter. *Reception characteristic:* the field is approximately proportional to D^{-1} for all wavelengths.

b) *Propagation via transition layers:* the field originates along a ray reflected (usually partially) at a layer in the troposphere where the refractive index varies abnormally strongly with height. *Occurrence:* particularly when a temperature in-

version or a meteorological front gives rise to the required transition layer. *Reception characteristic:* over a limited region, the field decreases more slowly than corresponds to a factor D^{-1}, and increases somewhat with the wavelength.

c) *Propagation as in a waveguide:* the propagation is brought about by multiple reflection of many rays between the earth's surface and the top of the "waveguide". This upper boundary is a region of indefinite thickness, at which continuous refraction — resulting in reflection — takes place. *Occurrence:* only under exceptional conditions, when with increasing height, a rapid decrease in humidity is accompanied by a rapid rise in temperature. *Reception characteristic:* for very short wavelengths, a field is observable below the horizon that is much stronger than the normal field brought about by diffraction, see (e), and decreases with increasing wavelengths.

d) *Gradient reflections:* the field is produced by partial reflections at the infinite number of infinitely thin layers of constant refractive index which constitute the continuous variation of refractive index with height in the troposphere. *Occurrence:* always present, but not noticeable except when they contribute to the field below the horizon due to diffraction. *Reception characteristic:* the decrease in the diffraction field just under the horizon is less pronounced than usual.

e) *Diffraction:* the field is produced by deflections of the rays towards a region beyond the horizon unattainable by geometrical-optical rays. *Occurrence:* always present, but contributing a significant field only at moderate distances beyond the horizon. *Reception characteristic:* the field decreases exponentially with distance, without being subject to pronounced fluctuations in time.

f) *Scattering:* the field is caused by scattering at small inhomogeneities that are present throughout the troposphere. *Occurrence:* always present, but only noticeable at great distances, where the other modes of propagation no longer contribute significantly to the field strength. *Reception characteristic:* the field at great distances decreases more slowly than exponentially, and is also very sensitive to weather conditions.

Summary. This second part of a survey of radio propagation in the troposphere deals with scattering. Inhomogeneities (eddies) are discussed, which occur in all sizes and are subject to turbulent motion. The smallest eddies that are at all likely to occur (micro-eddies) are especially important for scattering. In order to evaluate the scattered field, the field due to the scattering of one volume element is first considered. An interpretation then appears to be possible in accordance with Bragg's well known law. The total scattered field at a receiver is then found by integration over all active volume elements. The mathematical difficulties are minimized by assuming that the transmitter and receiver are beamed on one another. If a mathematically tractable function is now chosen for the volume fluctuations of the refractive index, numerical results can be obtained. Two such functions are mentioned together with the corresponding scattered fields. Observational data are next considered from the point of view of scattering, including observations of fading. The article ends with a recapitulation of the six idealized ways by which radio waves can be propagated in the troposphere, and gives their special characteristics.

PHOTOMETRY AT LOW LUMINANCE LEVELS

by W. de GROOT.

535.24

The establishment, by the Commission Internationale de l'Éclairage (Stockholm 1951), of a relative efficiency function for the dark-adapted eye (scotopic efficiency function) and the discussions held on that occasion, are once more drawing attention to the problem of photometry at low levels of illumination and corresponding low luminances. A survey of this matter is given below, which may be considered as a supplement to earlier articles on the same subject).*

Photometry

If a flat surface, for example a diffusely reflecting white screen, is exposed to radiation with wavelength(s) between 0.4 and 0.7 μ, an observer will obtain from it an impression of light — provided the power radiated on to the screen is not too low — and he will be able to form a judgment as to the brightness of the screen. This subjective "brightness impression" depends on a number of conditions. The principal ones are:

1) the power of the radiation emitted by the screen per unit area and the spectral distribution of the radiation,
2) the brightness of adjoining parts of the field of view and the state of adaptation of the eye,
3) the part of the retina involved in the observation.

With (1) we shall deal more in detail afterwards. As an example of the influence of the factors mentioned under (2) the appearance of the moon at night and in the daytime may be considered. If we look at the moon against a clear sky in daylight, we see it as a whitish patch on a blue background. After sunset the same moon appears to us as a glaring white disk with black surroundings. Nevertheless, the radiation which the moon sends to our eye is the same in both cases; only the brightness of the surrounding sky has diminished [1]).

An example of (3) is the well-known phenomenon of the disappearance of a small and feeble light at which we stare in the dark. It is caused by the fact that, in the dark, the central part of the retina (the fovea) is much less sensitive to light than the surrounding part (the parafovea) and the rest of the retina (the periphery). In photometry, the observer makes himself independent of this variability of the brightness impression, by comparing only fields of approximately equal brightness. In instruments constructed for the purpose (photometers), the observer sees two adjacent fields of view, one illuminated by the unknown source and the other — used for comparison — by a source of adjustable and known intensity.

In view of what has been mentioned under 2) and 3), care has to be taken that both fields of view are in the same condition as regards their surroundings and the part of the retina involved in the observation. The two fields of view may, for example, be the two halves of a circular field (*fig. 1*). The surroundings may be either dark or they may have a certain brightness, for instance, about equal to that of the circular field. The viewing angle of the photometric field will be chosen according to circumstances. With high brightnesses, at the level of illumination which makes colour vision possible, an angle of 2° is usually chosen, so that only the fovea and its immediate neighbourhood take part in the observation. With low brightnesses, having regard to the low sensitivity of the fovea, a larger angle (e.g. 6° to 10°) is chosen, so that the influence of the fovea may be disregarded and the observation is effected mainly by the parafovea.

Fig. 1. Photometer field (angular diameter 2° in photopic region, 6°-10° in scotopic region).

Influence of intensity and spectral distribution on the judgement of brightness

The influence of the conditions mentioned under (1), above, will now be considered, beginning with the case of high brightnesses (luminance > 3 cd/m^2).

*) See P. J. Bouma, Philips tech. Rev. **1**, 102-106, 142-146, 1936; G. Heller, Philips tech. Rev. **1**, 120-125, 1936; **5**, 1-5, 1940; W. de Groot, Philips tech. Rev. **10**, 150-153, 1948.

[1]) The radiation which we receive from the moon in the daytime is actually slightly greater, since the light of the sky itself is superimposed on that of the moon proper.

High brightness levels

At high brightness levels, the observer obtains impressions of colour and brightness simultaneously. It is difficult to separate the colour impression from the brightness impression. When comparing the two fields in a photometer, one illuminated by monochromatic green light, the other by monochromatic red light, it is difficult to state whether the brightness of the two patches is the same, although after some practice, judgement in this respect is improved.

In such cases, a step-by-step method is often used. The green colour is first compared to another colour, somewhat nearer (say yellow-green); this is compared to a colour which is still more yellowish, and so on until red is reached. This is convenient because the adjustment to equal brightness is much easier if the colour difference is small.

If, in this way, the whole visible spectrum is run through, it will be seen that the ratio of powers P_λ, to be radiated per unit area and solid angle in the direction of the eye, to ensure equal impressions of brightness at each wavelength, are different for each pair of wavelengths. For a certain wavelength (λ_m) in the neighbourhood of 555 mμ (5550 Å) the power is a minimum. The ratio $P_{\lambda m}/P_\lambda = v(\lambda)$ is called the relative efficiency of the observer at the wavelength λ. For $\lambda = \lambda_m$, $v(\lambda) = 1$, for other wavelengths $v(\lambda) < 1$.

If the experiment is repeated with the same observer at a still higher brightness level, the same value of λ_m and the same function $v(\lambda)$ are found. Thus in the region of high brightness levels, $v(\lambda)$ is independent of the level of illumination. If, on the other hand, the experiment is made by different observers, different functions $v(\lambda)$ and different values of λ_m will generally be found. The CIE (or Commission Internationale de l'Éclairage) therefore standardised, in 1924, an efficiency function which is considered to be valid for a "normal observer". We shall denote this function by $V_p(\lambda)$, because the region of high brightnesses is also called the *photopic* region [2]). The corresponding value of λ_m we shall denote by λ_p.

Unless the contrary is indicated, we shall suppose in the following that the observer is a "normal observer".

Suppose that the radiation to be investigated is not monochromatic but is provided by a source with spectral distribution $P'(\lambda)$. This means that the power [3]) (per m² and per unit solid angle) radiated in the direction of the eye in the interval λ, $\lambda + d\lambda$ is $P'(\lambda)d\lambda$.

If the brightness impression in the photometer is equal to that of a monochromatic power P_m at wavelength λ_m, experiment shows that

$$P_m = \int P'(\lambda)\ V_p(\lambda)\ d\lambda. \quad . \quad . \quad . \quad . \quad (1)$$

The expression on the right-hand side of equation (1), multiplied by a suitable factor, may also serve as a measure of the brightness. The usual unit is the candela per m² (cd/m²) or nit. If the brightness is defined in this way (and preferably measured in this unit) this quantity is given the name luminance (L). The unit cd/m² is so chosen that a perfect radiator ("black body") at the freezing temperature of platinum ($\approx$2042 °K) has a luminance of 6×10^5 cd/m².

From the known spectral distribution of the perfect radiator it is possible to calculate the value of the constant (K_p) occurring in the formula

$$L = K_p \int P'(\lambda)\, V(\lambda)\, d\lambda \quad . \quad . \quad . \quad . \quad . \quad (2)$$

The value [4]) of K_p is about 650 lm/W.

Low brightness levels

If the experiments just described are repeated with luminances less than 3 cd/m² the wavelength λ_m is found to be displaced to shorter values. This displacement depends on the part of the retina involved in the observation [5]). For the fovea the displacement is small, but at decreasing intensity the ability of the fovea to detect light soon ceases. As regards the periphery, already at high luminances the wavelength λ_m does not coincide with the value λ_p characteristic of the fovea. At decreasing intensity, λ_m shifts to still lower values. If the

---

[2]) From the Greek photos = light, and opsis = vision, in contrast with scotopic, from the Greek skotos = darkness.

[3]) In the literature, the symbol E(λ) is often used. We prefer the letter P because the quantity under consideration is a power and not an energy, and provide it with a dash to keep the reader aware of the fact that $P'(\lambda)$ has not the dimension of a power but of a power per unit of (wave) length, and thus has the character of a derivative with respect to λ.

[4]) If 3.74×10^{-16} W.m² and 14385 μ. °K are taken as the values of the constants $c_1 = 2\pi hc^2$ (radiation over a solid angle of 2π) and $c_2 = ch/k$ respectively, in Planck's radiation formula, and if the freezing temperature of platinum is based on optical pyrometry, taking 1336 °K as the melting temperature of gold, one finds $K_p = 683$ lm/W (see Philips tech. Rev. **10**, 150-153, 1948). This value is higher than the value found experimentally (about 630 lm/W). This points to the fact that the real melting temperature of gold is probably higher than the internationally accepted value (between 1340 and 1342 instead of 1336 °K; see F. Henning, Temperaturmessung, Barth, Leipzig 1951, pp. 140 and 256, W. de Groot, Physica **16**, 419, 1950).

[5]) See W. D. Wright, Researches on normal and defective colour vision. Kimpton, London 1946, pp. 73-87.

experiment is arranged in such a way that at high luminances the fovea alone takes part in the observation and at lower brightnesses the parafovea is allowed to take part in a prescribed way (e.g. by a gradual increase of the visual angle), it is permissible to say that the value of λ_m gradually shifts from the photopic value λ_p to lower values. At the lowest perceptible brightnesses, this shift comes to a standstill. The wavelength λ_m has then assumed the value 507mμ, which we shall denote by λ_s, because the region of lowest perceptible brightnesses is also called the *scotopic* region [2]).

With a special experimental arrangement, it is possible to define unique values of λ_m, and of an efficiency function $V(\lambda_m,\lambda)$, corresponding to each value of P_m. Conversely, a unique value of P_m may be found corresponding to each value of λ_m, between λ_s and λ_p. The form of the function $V(\lambda_m,\lambda)$ in which λ_m occurs as a parameter, depends on λ_m. If a field with spectral distribution $P'(\lambda)$ appears equally bright as a field radiating monochromatic light of wavelength λ_m with a power P_m, experiment shows [6]) that

$$P_m = \int P'(\lambda)\ V(\lambda_m,\lambda)\ d\lambda.$$

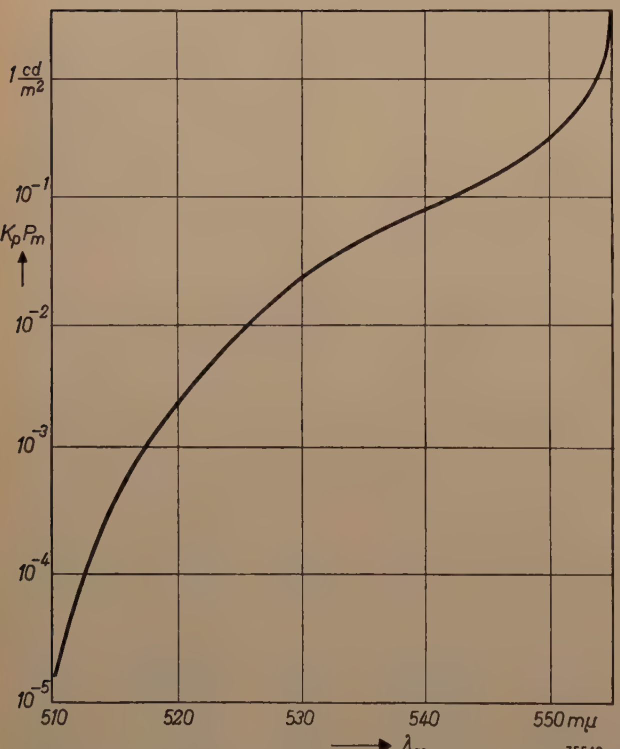


Fig. 2. P_m as a function of λ_m (Weaver). The quantity K_pP_m (expressed in cd/m²) is plotted on a logarithmic scale.

6) P. J. Bouma, Proc. Kon. Acad. Wet. Amsterdam, **38**, 33-35, 1935, P. J. Bouma, Physical aspects of colour, Ch. II (Philips technical Library, 1947). See also W. de Groot, Appl. sci. Res. **B2**, 131-148, 1951, J. Vorobeitchik, Bull. Soc. Belge. Electr. **68**, (1952), No. 2.

The connection between λ_m and P_m was first investigated in 1890 by König [7]).

Up till now the functions $P_m(\lambda_m)$ and $V(\lambda_m,\lambda)$ have not been standardised, as agreement has not

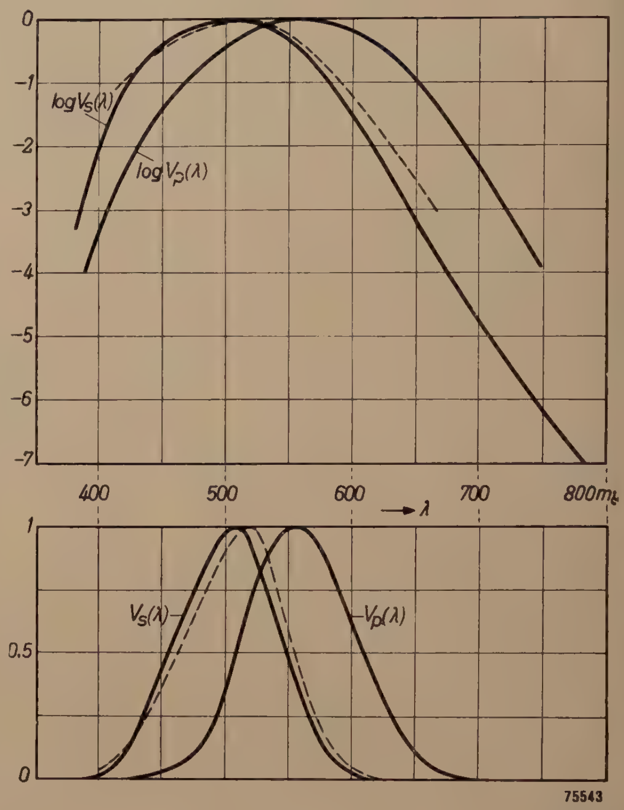


Fig. 3. The functions $V_p(\lambda)$ (CIE 1924) and $V_s(\lambda)$ (CIE 1951) plotted against λ. In the upper graph the ordinate is logarithmic, and in the lower graph, linear. The broken line corresponds to Weaver's $V_s(\lambda)$-function.

yet been reached on the way in which the experiment should be arranged. In 1949 Weaver [8]) published a series of tables for the functions $V(\lambda_m,\lambda)$, which may serve as a basis for calculations. In 1951, at the CIE conference at Stockholm, a function $V_s(\lambda)$ was standardised [9]), valid for young eyes, in the region of very low brightnesses. In *fig. 2* the connection between P_m and λ_m according to Weaver is shown, and in *fig. 3* the functions $V_p(\lambda)$ and $V_s(\lambda)$, as given by the CIE.

Photometry at low brightnesses

What are the consequences of all this for photometry at low brightness levels? Suppose that in a photometer, two fields of view are observed,

7) See P. J. Bouma, Philips tech. Rev. **1**, 142-146, 1936. P. G. Nutting, Bull. Bur. Stand. **7**, 236-238, 1911.
8) K. S. Weaver, J. Opt. Soc. Amer. **39**, 278-291, 1949.
9) Proc. CIE (Stockholm 1951), part III, pp. 33-39. For the sake of clarity in the present article, the symbols K_m, K_m', $V\lambda$, $V\lambda'$, as used by the CIE, are replaced by K_p, K_s, $V_p(\lambda)$ and $V_s(\lambda)$.

radiating light of wavelengths λ_1 and λ_2 respectively, with corresponding powers P_1 and P_2 such that the observations are made in the photopic region. When luminances are equal, we have

$$K_p P_1 V_p(\lambda_1) = K_p P_2 V_p(\lambda_2).$$

If we now decrease both powers by multiplying them by the same factor (<1) — this may be done, for example, by inserting a neutral filter — so that we are no more in the photopic region, both fields will, according to definition, still have the same luminance, but to the eye they will no longer appear equally bright. Conversely, two fields appearing equally bright will no longer have the same luminance if we are not in the photopic region.

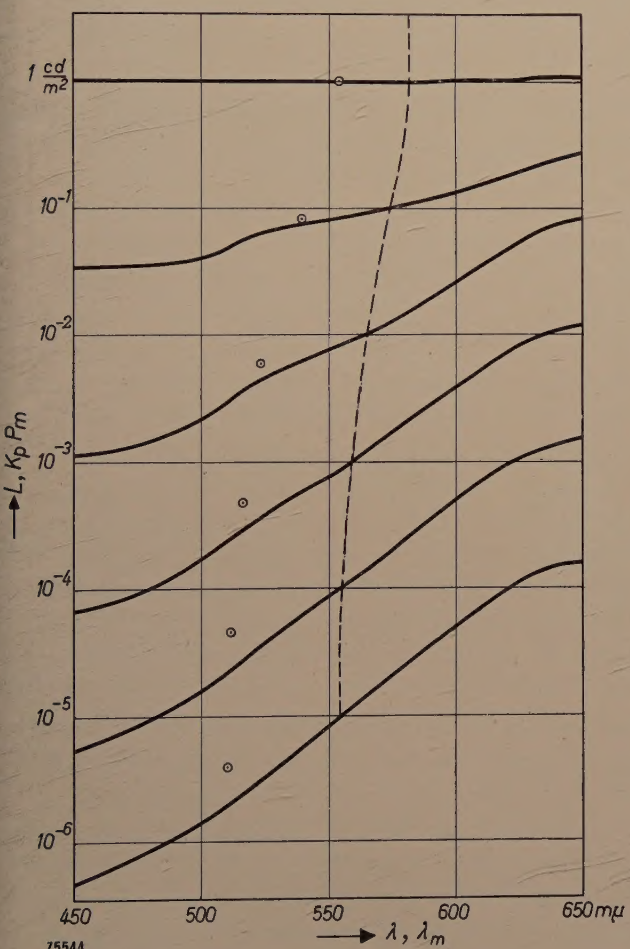


Fig. 4. Lines of equal "subjective brightness" for monochromatic radiations, according to Bouma. Ordinates (logarithmic scale) are luminances L in cd/m². In the same diagram, values of $K_p P_m$ (see fig. 3) are plotted against λ_m. The broken line joins the wavelengths for which the luminance is equal to that given by a source of $T_c = 2042$ °K, appearing equally bright.

In *fig. 4*, which is due to Bouma, the luminances L of monochromatic fields appearing equally bright, are given on a logarithmic scale. Corresponding values are joined by a curve (full lines), each curve representing a certain brightness level. The values of $K_p P_m$ and λ_m corresponding to each curve are also indicated.

The function plotted against λ is

$$\log\,[K_p P_\lambda V_p(\lambda)] = \log\,[K_p P_m V_p(\lambda)/V(\lambda_m,\lambda)]\,.$$

In the photopic region $\lambda_m = \lambda_p$ and $V(\lambda_m,\lambda) \equiv V_p(\lambda)$. Therefore, in this region, the value is constant and equals $\log\,[K_p P_m]$. In the region of low brightnesses, $\lambda_m < \lambda_p$ and $V_p(\lambda_m) < 1$, whereas $V(\lambda_m,\lambda_m) \equiv 1$. Therefore, the point with coordinates λ_m and $\log\,(K_p P_m)$ is above the corresponding curve.

The points of any one curve in fig. 4 correspond, as already indicated, to fields appearing equally bright. The curves are plotted in order of increasing apparent brightness.

We now put the question: is it possible to define a "measure" for this apparent brightness? To each curve corresponds a value of P_m. The higher the value, the higher the curve will be in the diagram. It is natural to use P_m as a measure of apparent brightness. We may also use $K_p P_m$, i.e. the luminance of the field radiating monochromatic light of wavelength λ_m with power P_m, which appears equally bright. But it is evident that an arbitrary monotonic function $f(P_m)$ (i.e. a function which always increases with increasing P_m) also satisfies the requirements [10]).

Bouma, following König, chose the luminance of a field appearing equally bright and radiating monochromatic light of wavelength $\lambda = 535$ mμ as a measure of what he called "subjective brightness". This definition is unambiguous, but it is not well suited for photometric practice, because a source radiating this wavelength is not readily available.

It is, however, not necessary to restrict oneself to monochromatic sources, because the choice of the function $f(P_m)$ is arbitrary within wide limits. A definition which has found practical application uses as a light source an incandescent lamp of colour temperature 2360 °K. This, of course, is the type of lamp used in most photometers. The subjective brightness is now made equal to the luminance of a field appearing equally bright and radiating light with a relative spectral distribution equal to that of a black body of the said temperature. The advantage of this definition is that now "subjective brightnesses" can be measured using an ordinary photometer.

The colour temperature of 2360 °K is rather arbitrary, and another choice might be considered.

[10]) P. J. Bouma, Physica 8, 413-424, 1941.

At the CIE conference 1951 this question was discussed and agreement was attained on the desirability — at least from a theoretical point of view — of using the same light source, for the measurement of the "subjective brightness" as used to fix the unit of light intensity (candela) at high luminances, viz. a perfect radiator with a temperature of about 2042 °K.

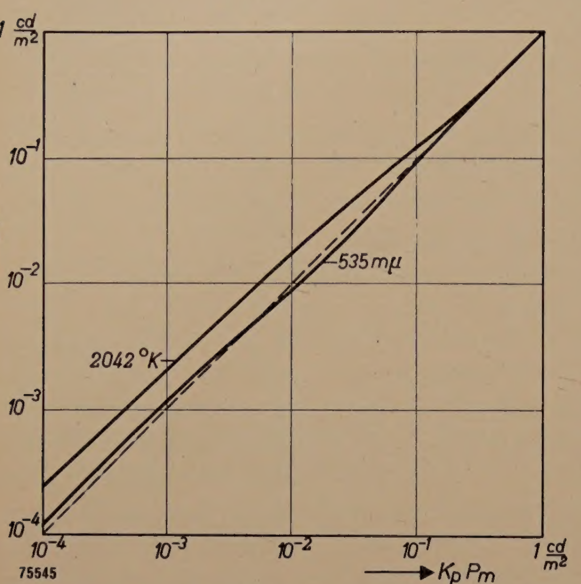


Fig. 5. Luminance of photometer fields, illuminated by two different sources (λ = 535 mμ, T_c = 2042 °K), plotted against K_pP_m (both logarithmic). The broken line would correspond to K_pP_m equal to L. The scotopic region is the region where $K_pP_m < 10^{-4}$ cd/m², and the photopic region where $K_pP_m > 3$ cd/m².

In *fig. 5* the luminances of two radiations (monochromatic, λ = 535 mμ, and black body with T = 2042 °K) are plotted against K_pP_m (both on a logarithmic scale). This figure therefore shows the function f(P_m), both for the definition of König-Bouma and for that of the CIE. It is seen that at high brightnesses the two functions are identical, and at the lowest brightnesses (in the scotopic region) K_pP_m and the "subjective brightness" are again proportional to each other. Their ratio [11]), however, depends on the choice of f(P_m). The latter fact simplifies the case in the scotopic region with which we are often concerned. In this region the "subjective brightness" is represented by the expression

$$K_s \int P'(\lambda)\, V_s(\lambda)\, d\lambda,$$

which formula is analogous to formula (2) for the luminance, with the difference that $V_p(\lambda)$ is replaced by $V_s(\lambda)$ and the factor K_p by K_s. The value of K_s depends on the definition of f(P_m), and therefore on the choice of the standard source. For the CIE source (T_c = 2042 °K), taking the CIE values for $V_s(\lambda)$, K_s = 1746 lm/W.

The question arises, whether a light source exists for which, in the entire region, the subjective brightness is simply proportional to P_m, in which case $f(P_m)$ would be represented by a straight line in fig. 5. Such a light source does not exist, as was pointed out by Bouma. If a source with colour temperature T_c = 7500 °K were chosen [12]), K_s would be equal to K_p.

Another way of arriving at the same result [13]) would be to choose a monochromatic source, for which $V_s(\lambda) = V_p(\lambda)$, i.e. a source with wavelength λ = 528 mμ. In both cases, however, f(P_m) is not proportional to P_m in the intermediate region. The deviation from the straight line is greater for the curve corresponding to λ = 528 mμ than for that corresponding to T_c = 7500 °K.

The proposal of the CIE

The choice of the CIE reduces photometry at low brightness levels to a simple question (at least in theory). The comparison field in the photometer is to be illuminated with light of the candela-source, which is to be attenuated by a neutral attenuator such as a rotating sectored disc, a neutral filter or a diaphragm. The attenuation is to be regulated in such a way that the comparison field, illuminated by the standard source, appears equally bright as the field to be measured. The result is obtained in units cd/m², as in the case of high luminances.

A difficulty arises because the sensitivity curve of the parafovea differs from that of the fovea, as explained above. Hence, in reality the question is not so simple, owing to the great difference between the spectral distribution of most light sources and that of the source of colour temperature T_c = 2042° K. Even if the colour difference is no hindrance in itself (as is the case with parafoveal or peripheral observation in the intermediate and in the scotopic region), measures have to be taken to ensure that the fovea is not involved in the observation [14]).

In practice, therefore, a photometer lamp with a higher colour temperature (2360 °K or more) is chosen. The measured luminance will then have to be corrected, in order to reduce it to the desired luminance of a field of 2042 °K appearing equally bright. In the scotopic region this may be done simply by multiplying by a constant factor.

One question remains: in which unit is the "subjective brightness" to be expressed? If the

[11]) See also W. D. Wright, Photometry and the eye, Hatton, London 1949.

[12]) W. de Groot, see note [6]).

[13]) P. J. Bouma, see note [10]).

[14]) A case in which these conditions were not observed is described in Proc. CIE 1951 (III), Gen. Paper D: A. M. Kruithof and W. de Groot, Photometry at very low brilliance levels.

recommendations of the CIE are followed, a result expressed in cd/m² is obtained. It is then possible, however, that for a certain lamp, a different number of lm/W is obtained depending on whether the illumination belongs to the scotopic or the photopic region (as may be the case with the illumination of roads).

In general, the scotopic radiation equivalent

$$k_s = K_s \frac{\int P'(\lambda)\, V_s(\lambda)\, d\lambda}{\int P'(\lambda)\, d\lambda} \text{ lm/W (s)}, \quad . \; . \; . \quad (3)$$

differs appreciably from the photopic equivalent

$$k_p = K_p \frac{\int P'(\lambda)\, V_p(\lambda)\, d\lambda}{\int P'(\lambda)\, d\lambda} \text{ lm/W (p)}. \quad . \; . \; . \quad (4)$$

Omission of the addition "scotopic" (s) or "photopic" (p) to the symbol lm/W might give rise to misleading statements.

Table I. Ratio of radiation equivalent (lm/w) in the scotopic region (k_s) to that in the photopic region (k_p), for various light sources [15]).

Monochromatic light		Perfect radiators	
λ(mμ)	k_s/k_p	T_F(°K)	k_s/k_p
450	21 (30)	2042	**1.00**
500	7 (7.7)	2360	1.2
550	1.3 (1.2)	2850	1.4
600	0.2 (0.13)	5500	2.15
650	0.07 (0.02)	7500	2.4
sodium	0.3 (0.23)	sunlight	2.0
mercury	1.1	daylight	2.3

[15]) The figures are calculated from Weaver's data. The figures in brackets in the first column are from the CIE data. The composition of the mercury light source is: λ 405 mμ (12%), λ 408 mμ (2%), λ 436 mμ (22%), λ 546 mμ (29%), λ 578 mμ (35%).

The above-mentioned expressions (3) and (4) have the same value only if $P'(\lambda)$ represents the spectral distribution of a source with $T_c = 2042$ °K (the calculation of K_s from K_p was based on this equality).

In Table I the ratio k_s/k_p is given for a number of light sources. It is found that for light sources with a continuous spectrum (incandescent lamps) the ratio is not greater than 2 or 2.5. It may reasonably be supposed that the same holds for fluorescent lamps. It is remarkable, however, that for monochromatic light, especially for wavelengths near the ends of the visible spectrum, the values of k_s and k_p may differ by a factor of more than 10.

Summary. In photometry the observer makes himself independent of the variance of brightness impression by always comparing two patches of light which produce about equal brightness impressions. In the region of high brightnesses an efficiency function of the human eye has been standardised (CIE 1924). The definition of luminance is based on this function. The unit of luminance is the cd/m² or "nit". At low luminances (< 3cd/m²) equality of luminance does not correspond to equality of brightness impression. Following Bouma, "subjective brightness" is defined as the luminance of the comparison field, illuminated by a special source, which in the photometer appears equally bright with the field to be measured. For this source, Bouma chose monochromatic light (λ = 535 mμ). Another often recommended choice is a source of colour temperature T_c = 2360 °K. In 1951 the CIE recommended a source with T_c = 2042 °K (black body at the freezing temperature of platimun, on which the definition of the candela is also based). At the same time, an efficiency curve for the region of lowest brightnesses (scotopic region) was standardised. It is now possible to express the "subjective brightness" in cd/m², and to use the units lumen, lux etc. also in this region. In order to avoid misunderstanding, however, it is advisable to mark these units with the suffix: scotopic or photopic. In scotopic photometry use can be made of other standard sources (e.g. T_c = 2360 °K, T_c = 7500 °K), but in this case the measured luminances must be multiplied by a correction factor in order to reduce them to the luminance of an equally bright field, illuminated by the candela source.

BOOK REVIEW

Artificial light and photography, by G. D. Rieck and L. H. Verbeek, pp. 347, 180 illustrations, 52 plates (4 in colour), 10 tables. — Philips Technical Library — This book can be ordered through your technical bookseller.

On opening this book, the first impression is one of mathematical formulae set out against a background of exquisite photographic reproductions. This is only superficially an incongruous combination: in fact, it serves to demonstrate the two-fold nature of photography, being an art and at the same time a science, or better, a technique based on scientific principles.

In the handling of this technique and its varied apparatus (e.g. cameras, developing methods, enlarging and projection equipment and their associated light sources) the serious photographer will want to know something of the fundamentals of the subject and the factors which influence his results. He will find the present book to be a rich source of information. The authors have succeeded in producing a comprehensive and often enthralling picture of the scientific background of their subject. The book is suitable both for the professional photographer with a considerable photographic background and for the intelligent and technically interested amateur. Some of the more specialized matters are dealt with in detail in small print and a number of literature references are given. On the other hand, readers who seek practical information can omit the theoretical parts, which are chiefly concentrated in chapters II and III (Light and lighting, and sensitive materials and their reaction to light, together about 100 pages), and enjoy chapters V, VI and VII (exposure using filament and gas-discharge lamps, flash exposures and the use of artificial light for projection, reproduction and processing — together 120 pages). Many practical hints, useful to both professional and amateur, are given. In between these sections of the book, Chapter IV (80 pages) gives an extensive description of the light sources themselves. Finally, Chapter VIII briefly discusses some special applications of artificial light, e.g. for infra-red and ultra-violet photography, microphotography etc.

The contents of the book are well balanced. For example, considerable attention (more than usual in books of this nature) has been given to the construction and characteristics of the incandescent lamp and the flashbulb, in view of their practical importance. Some sections of the book deserve separate mention, because of their original presentation of certain practical problems: the discussion on the combination of sunlight and flashlight (p. 239), the "shadowing" of objects (p. 37), and the colour rendering of monochrome and colour materials (p. 66) are especially noteworthy. As a handbook its value is greatly enhanced by the tables of practical data and by the two extensive indexes (authors and subjects).

Distributed througout the book are about fifty large reproductions (some in colour) of pictures taken with artificial light. Small-scale reproductions of these pictures are printed at the end of the book together with their technical data, in particular the lighting methods used. This "concise course in artistic photography" may well induce the serious amateur to aim at emulating these results in his own work.

S. GRADSTEIN.